1979

A Behavior Modification Intervention Program in a Residential Treatment Center for Emotionally Disturbed Boys

Richard Alan Stern
Loyola University Chicago

A BEHAVIOR MODIFICATION INTERVENTION PROGRAM IN A RESIDENTIAL TREATMENT CENTER FOR EMOTIONALLY DISTURBED BOYS

by

Richard A. Stern

A Dissertation Submitted to the Faculty of the Graduate School of Loyola University of Chicago in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

May

1979

ACKNOWLEDGEMENTS

I would like to offer special appreciation to Dr. Frank Kobler for his continued support, encouragement and guidance throughout the study. Special appreciation is also given to Dr. Al De Wolfe for the time he spent assisting me in the statistical areas of the project, and to Dr. Roderick Pugh for his encouragement, supportive attitude, and perceptive feedback.

Special acknowledgement is also given to Gene B. Meier, Executive Director of Lawrence Hall School for Boys for making available the resources of the Agency for this project. Additionally, I would like to acknowledge the hours of work above and beyond the call of duty put in by the child care staff at Lawrence Hall. These dedicated professionals filled out many point sheets and rating forms and did so in a manner that exemplifies their dedication to their chosen field.

VITA

The author, Richard Alan Stern, is the son of Richard Eisendrath Stern and the late Babette (Kaplan) Stern. He was born November 18, 1947 in Chicago, Illinois.

Richard Stern's elementary and high school education were obtained at the University of Chicago Laboratory Schools where he graduated in 1964.

In May, 1968, he received his Bachelor of Arts Degree from the University of Michigan at Ann Arbor with a Major in Psychology. While attending the University, he was an Associate Editor of the campus newspaper.

In January, 1972, he received his Masters Degree in Psychology from Temple University in Philadelphia.

In addition to his Doctoral studies at Loyola, he has been associated with Forest Hospital, the Jewish Children's Bureau of Chicago, and Lawrence Hall School for Boys. He is also founder and director of Camp Sullivan, a therapeutic residential camping facility for children with adjustment and learning disabilities.

TABLE OF CONTENTS

TABLE OF CONTENTS.--Continued

LIST OF TABLES

LIST OF TABLES.--Continued

I. INTRODUCTION: STATEMENT OF THE PROBLEM

The present study is a description and analysis of a behavior modification intervention experiment in a residential treatment center for emotionally disturbed boys.

There is a great need for well-designed scientific experimentation in the area of residential treatment of emotionally disturbed children. Residential treatment of children is growing faster as a field than the body of knowledge underlying it. State and Federal funding sources expend large amounts of money on programs that are presumed to have beneficial effects on the children enrolled in them. Yet very little effort goes into the evaluation of the effectiveness of these programs.

Another recent development in the field of residential treatment is the increasing utilization of behavior modification techniques in various forms ranging from point systems utilized with individual children to token economies around which entire programs are based. Again, the assumption is being made that these are effective therapeutic techniques although the underlying body of knowledge in support of these assumptions, specifically in regards to residential treatment, is inadquate.

Krasner (1971), however, has pointed out that behavior modification lends itself to meeting accountability demands. Three of its major characteristics all seem germane to the present study: (1) concepts are stated in such a manner that they can be tested experimentally; (2) an explicit strategy of therapy exists; and (3) the goals of the modification procedure can be determined by an initial assessment of the problem behaviors.

The present study attempts to contribute to the body of knowledge relating behavior modification to residential treatment. It attempts to define a methodology that is original and yet useful for other studies of this type. A significant contribution made in this study is the development and analysis of a replicable methodology which meets many criteria for research. Adding to its significance is that it takes into account ethical implications that stem from issues involving harm done to an untreated control group. The methodology also allows for the development and investigation of new behavior modification programs which can be implemented and compared in effectiveness with already existing programs being carried out with untreated control subjects. If the particular behavior modification program is found to be beneficial then it can be used for (former) control subjects also.

Additionally, particular segments or highlights of a program that are effective can be statistically analyzed

and then utilized with specific sub-groups of subjects. Although research of this type focuses on only one aspect of a total treatment program, it should also be kept in mind that more global and general studies have failed to prove that residential treatment is any more effective than no treatment at all. Even the few studies that have indicated limited outcome success after treatment have failed to be able to correlate success to any particular treatment variables.

Breaking the residential treatment process into small units for analysis will assist in beginning to determine what if any components do contribute to effective treatment when it happens.

II. REVIEW OF RELATED LITERATURE

According to Pappenfort, Dinwoodie, and Kilpatrick (1968), as of 1968 there were nearly 150,000 children residing in approximately 2,500 child care institutions in the United States. Of these 2,500, only 200 were judged to be providing treatment. However, 110,000 of the 150,000 institutionalized children were judged to be emotionally disturbed. Of those in need of treatment, 14,000 or 13 percent were judged to be getting it.

This survey will focus on some of the characteristics of and research in regard to those institutions that are providing treatment. Herstein (1975) has pointed out that private residential centers are becoming increasingly dependent on public funds for support, and that accountability in terms of results is therefore of greater significance than it has ever been in the past. Yet, as will be shown later, when one surveys the field in terms of existing research, many deficiencies appear to exist. The majority of the studies have been outcome or follow-up studies that did not employ control groups or define a rigorous methodology.

The first section of this study will focus on characteristics of and current issues in residential treatment. The remainder of the paper will consist of a

critical review of the literature and a discussion of the problems of research methodology as it is applied to residential treatment.

Characteristics of Residential Treatment

Most authors do not deal with the issue of defining residential treatment. However, Birnbach (1971) has provided a definition:

> A residential treatment center for children is a total institution in which all aspects of the child's life take place in the limited arena of the institution's own grounds. Its function is to offer the individual child a number of related experiences that are designed to help him regain some control over his life and the circumstances surrounding it (p. 177).

Mayer (1955) broadens this definition by asserting that residential treatment is not an entity in itself but a stage in a total treatment process, which includes pre-institutional as well as post-institutional care.

Inglis (1964) describes a philosophy of treatment that seems to apply to many contemporary settings. This philosophy, developed out of six years experience at High Meadows, a state institution serving 36 boys in New York, ages 6-15:

> We can offer a manageable, relatively protected reality situation in which all our staff assist the child in gradually learning to cope with it . . . the treatment institution must combine its resources to restore, as much as possible, the non-functioning parts of the child's personality . . . only as the uncontrolled behavior is brought under some form of control is the child able to begin to learn and do, and look at himself (p. 279).

The emphasis on reality as well as the external control of behavior has been emphasized by other authors as well (Easson, 1969; Glasser, 1975; Trieschman, Wittaker & Brendtro, 1969).

One issue that is not given any attention in the literature is the dichotomy between hospital inpatient treatment as "residential treatment," and the more traditional child welfare social agency as "residential treatment." Clearly there are significant differences between the two settings in terms of budget, treatment modality, and qualifications of treatment personnel. Though some authors discuss findings with reference primarily to hospital settings, and others with reference primarily to agency settings, the implications for research of the differences between types of settings are not clearly demarcated. Thus one finds the literature consisting of a body of articles describing work done at hospitals, and a somewhat larger body describing studies done in child welfare settings; yet, all are classed under the rubric of "residential treatment."

In order to best understand some of the issues relating to the research question in residential treatment, it is useful to consider a historical perspective. Browne (1963) reported that most of today's non-hospital residential facilities emerged from sectarian institutions

whose original goals were shelter, care, and training.

> The development in the past four decades of supportive services (e.g., family counseling) . . . diminished the old demand for long-term custodial care. The increasing use of psychiatrists as consultants in custodial institutions helped to shift the emphasis from care to treatment (p. 231).

In other words, most residential centers evolved from orphanages and foundling homes which operated as low budget, "end of the line" institutions where primary consideration could not possibly be given to treatment and evaluation. However, in the past two decades the welfare agencies have opened their doors to emotionally disturbed and behaviorally disordered children and have had to turn more and more to public funding of one sort or another for support.

In the 1940's and 1950's those attempts made at going beyond custodial care into treatment were influenced by Aichhorn (1934), working primarily with Viennese delinquents, who advocated a closed system approach wherein the child in treatment had minimal contact with the outside community. It was hoped that the milieu would have a strong impact on the pathology of the child. In the 1950's the work of Bettelheim (1950) at the Orthogenic School in Chicago and Redl (1959) at Pioneer House in Detroit contributed significantly to the development of residential treatment, particularly through emphasis on the role of group living and the child care worker. Both Bettelheim

and Redl reported positive results of treatment and utilized primarily case history material to support their results. Both also operated primarily from a psychoanalytic framework applied to a milieu living arrangement.

What type of children are treated in contemporary residential settings? Often the catch-all phrase "emotionally disturbed" is used to describe a population, but without any attempt to define what this phrase really means. The lack of adequate description of subjects will be discussed later in terms of its relevance to methodological problems. Some authors have tried to describe the types of youngsters that are typically seen in residential treatment. Finkelstein (1974) categorized the three most common family constellations seen in new admissions. These are: (1) The child who is from an intact, but highly conflictual family; (2) the child who is from a one-parent family, or where the second parent is not a natural parent; or (3) the child who has virtually no family ties, and because of multiple traumas and failures generally winds up in a residential facility by the onset of adolescence. It was pointed out that many children who fall into the above categories do not end up in residential placement. Outpatient therapy and foster homes are often utilized prior to the more drastic and expensive step of placing the child in an institution.

Easson (1969) is of the opinion that only certain children are appropriate for residential placement. He presents a diagnostic framework that he has used primarily in a hospital setting. There are two major criteria for admission:

> 1) From the examination process it should be clearly demonstrated that the disturbed teenager lacks sufficient personal strength to control his own drives and impulses . . . a profound deficit in ego strength should be present before hospitalization is considered.
> 2) It should be shown that the teenager lacks the emotional capability necessary to form sufficiently strong meaningful relationships with people in his family and his culture . . . a profound deficit in relationship capability should be demonstrated before hospitalization is prescribed (p. 5).

Easson also comments on the dangers of hospitalizing children who are not really disturbed enough to require it. He suggests that this may produce permanent emotional handicaps. He points out that parents are sometimes eager to place a child in an institution as a means of punishment for acting out behavior. There is nothing wrong with punishing destructive behavior, but treatment agencies need to guard against being misused by parents in this manner.

Glickman (1957) differentiates between "closed" and "open" residential settings as the two primary structural modes of care. The closed setting more often provides a hospital type of treatment for severely disturbed or psychotic children. In contrast, the open setting, for less disturbed children, offers a diversified program

providing an opportunity for a corrective emotional experience. Of eight residential treatment centers surveyed in the Chicago area, only one has a closed unit available.[1]

The cost of residential treatment is quite high. The average per diem rate at the eight centers surveyed in Chicago in 1977 was $43.18 per boy.[2] The range was from $27.52 per day to $65.25 per day for the agency offering the closed unit program. These costs seem typical of residential treatment centers nationwide. Two hospitals were also surveyed. One reported a per diem of $160.00 and the other a per diem of $225.00. Both of these provided short-term closed unit residential care for children and adolescents.[3]

Browne (1963) has pointed out that

> in many institutional settings for children the words "therapy" and "treatment" have gained tremendous importance but they carry with them an aura of mystery. Just what "treatment" is understood to mean is persistently elusive. . . . There is vagueness regarding matters not only of definition, but also of formulation, prediction, and procedure (p. 73).

Adding support for this position, the Joint Commission on Mental Health of Children (1970) stated that

---

[1]Larkin Home.

[2]Maryville Academy, Lawrence Hall School, Chapin-Hall, Allendale School, Edison Park Home, Methodist Youth Services, Larkin Home, Mary Barthelme Homes.

[3]Forest Hospital, Northwestern Memorial Hospital.

> the present institutional arrangements for residential care of disturbed children are inadequate in many ways, requiring a critical appraisal of both assumptions and procedures (p. 42).

One issue that has become important in recent years involves the distinction between psychotherapeutic and so-called custodial functions in residential treatment. Several writers (Bettelheim, 1950; Birnbach, 1971; Browne, 1963; Herstein, 1977) have criticized this distinction as detrimental to the child and have argued for professionalization of the child care worker as an integral part of the treatment team. Browne (1963) feels that the role of child care worker or cottage parent is currently paradoxical and demoralizing, and that support and training are the solution to this problem.

Birnbach points out that child care is often viewed by other disciplines as a menial set of tasks designed to maintain the child for more significant treatment intervention elsewhere. If the child care worker accepts this role,

> it is more likely that the transactions between the child and the agency staff and between the individual and his small society cannot contribute to the overall objectives of both agency and client (p. 178).

Herstein has analyzed treatment centers in terms of those where there is a cohesive group process and those where conflict is latent but devastating.

> One source of latent staff conflict is the primacy given to individual psychotherapy . . . it is our

> hypothesis that unresolved latent staff conflict flourishes in those residential treatment settings whose model emphasizes the role of the expert, primacy of individual psychotherapy, dyadic relationship, and minimization of group relationships (p. 320).

If morale is low and conflicts are submerged rather than dealt with in the open, the child will suffer. Herstein sees the involvement of child care personnel in problem solving and decision making as essential to overcoming this source of dissension.

Follow-Up and Outcome Studies

The first area of the research literature to be surveyed consists of studies that come under the rubric of follow-up or outcome evaluations. Generally the purpose of such studies is to measure client characteristics before and at the end of treatment; also, frequently the studies attempt to ascertain progress one or more years after the conclusion of residential treatment.

Investigators seeking treatment variables which are associated with improvement or success have generally reached the conclusion that the single best predictor of success is the original strength or health of the subject (Davids, Ryan & Salvatore, 1968; Eisenberg, 1957; Garber, 1972; Herrera, Lifson, Hartmann & Solomon, 1974; Kane & Chambers, 1971; Levy, 1969; Warren, 1965).

However, these and other investigators have been more successful in relating follow-up results to variables

having to do with the environment provided after discharge. Several authors have demonstrated that the more support built into the aftercare experience, planned or not, the more successful is the outcome (Allerhand, Weber, & Haug, 1966; Novotney & Burstein, 1974; Persons, 1967; Taylor & Alpert, 1973).

Three of the studies reviewed did relate treatment results to the content of the treatment being provided (Coche & Thomas, 1975; Goldenberg, 1971; Persons, 1967).

Generally, when statistical results of treatment were reported, those cases judged to be successful were in the range of 33 to 60 percent, but successful cases were, as mentioned above, not usually found to be correlated to intervention strategies. Only Coche and Thomas (1975) reported a success rate of higher than 60 percent. One problem in relation to reporting success is that there is no clear-cut definition of what success actually is.

All of the fourteen studies reviewed in this section utilized differing methodologies and different measuring instruments. The only real overlap on instruments was the WISC and the psychiatric interview, each used by several investigators.

Of the fourteen studies, only two (Goldenberg, 1971; Persons, 1967) utilized a control group and only

Persons was able to assign subjects randomly. The implications of this are significant and will be discussed at the conclusion of this paper.

Follow-Up Only

In the first study to be reviewed, data was collected only with regard to post-hospital adaptation. Kane and Chambers (1961) viewed improvement as the patient's ability to cope with his post-discharge environment. They did a follow-up study involving twenty-four children, and collected data an average of seven years after the children had been discharged from residential treatment. The methodology consisted of interviews with parents and children and focused on such factors as attitudes toward the child, satisfaction with post-treatment arrangements, and dynamics of family relationships. Their results seemed to have been largely inconclusive and caused them to speculate that:

> since functional illness is a social phenomenon, improvement is also socially determined, and it is this network of complex interactions that makes it next to impossible to quantify improvement in any way that will yield a true picture. Outcome is related not only to original diagnosis, but also to a complex of attitudinal and social factors. In many cases one is left to conclude that the critical elements in improvement can never be isolated. Or sometimes improvement, like beauty, lies in the eye of the beholder (p. 1026).

Although these speculations are philosophically interesting, they are really more pessimistic than what other researchers have concluded. Most researchers seem to

maintain the position that existing studies have not adequately described the etiology of improvement, but that with more effort and vigorous methodology, such description is not impossible. It should also be pointed out that Kane and Chambers' study is the oldest of all the follow-up and outcome studies.

Studies of Antecedent Factors

Three studies collected information on the patients' personality or circumstances prior to, or at the onset of, treatment.

An analysis of 157 adolescent boys and girls was reported by Warren (1965) of the Bethlen and Maudsley Hospitals in London with respect to outcome six or more years after discharge. Originally, Warren studied case records of 204 adolescents discharged between 1949 and 1953. Diagnostically, the population was broken down into four major categories: Neurotic disorders, conduct disorders, mixed neurotic and conduct, and psychotic. Of the original 205 subjects, 157 were actually interviewed and included in the data analysis. The mean age at admission had been 14.5 for boys and 14.9 for girls. Roughly 25 percent had been hospitalized less than three months; 25 percent three to six months; and the remainder generally nine months to two years.

It was found that 33 percent of the neurotic and

25 percent of the mixed category had experienced further serious illness resulting in some form of residential treatment. Nearly all of the psychotic groups were still regarded as seriously ill. With regard to conduct disorders, most of them did not have further institutionalization but 50 percent had further incidence of anti-social behavior of a serious nature. It was also found that patients admitted for treatment initially when they were young fared much better than those admitted at an older age.

Warren's study was highly complex and he presented a mass of data generally obtained by the case history method and analyzed primarily according to his four diagnostic categories, which seem greatly oversimplified. He himself is cognizant of some of the problems of this study:

> The present study, while tracing and measuring as far as could be done the occurrence of psychiatric disturbances from early on, through adolescence into adulthood, shows very few significant relationships between them. The patients concerned were diverse and could for analysis be grouped only into broad diagnostic categories; nevertheless, some indications emerged of the likely outcome for further illness or anti-social behavior as between these categories, and so provide perhaps some pointers for further study (p. 158).

Levy (1969) studied 113 children who had been discharged from Southard School, a psychoanalytically oriented treatment center, between 1945 and 1960. Seventeen of his subjects had been treated in the school for less than six months; 79 from six months to three years; and 17 from three to ten years.

The original diagnoses were chronic brain syndrome (12 cases), psychotic (27), neurotic (21), personality disorder (13), neurotic behavior (24), and emotional maladjustment (16). It is not made clear in the study what the exact definition of some of these categories is.

Responses were collected from patients themselves, their families, and outside professional sources. The results were as follows: 34 were found to be leading normal lives, 24 had a "marginal adjustment," 6 were seriously ill but improving, 11 were chronically ill, 10 were in hospitals, 3 in prisons, and 5 deceased. Furthermore, 2 had Ph.D.'s, 3 had M.D.'s, 9 had M.A.'s, and 19 had B.A. degrees. Twenty-five had stable marriages, 12 were divorced, and the remainder had never married.

Levy reported that the combination of psychosis and low I.Q. carried a very grim prognosis with all of the subjects in this category doing poorly.

Levy's overall analysis of contributing variables yielded results similar to those of David's:

> Although further study will be necessary to specify exact characteristics, it appears that attractiveness and likeability, intelligence, verbal facility, an absence of overt aggressiveness and other well known "good patient" features are highly relevant (to successful adjustment), an observation distressingly bland (p. 1637).

Levy also analyzed data on 24 patients whose treatment was terminated by their families rather than by the

agency itself. He found that 58 percent of these patients were in the category of ordinary or marginal adjustment. He also found predictably that patients prematurely removed by well-adjusted families did much better than those removed by highly disturbed families.

Herrera et al. (1974) did a ten-year follow-up study of 55 young adults who had been hospitalized psychiatrically as adolescents. The results showed that only one-fifth of the total sample could be considered to be functioning well in society. Specifically, 11 subjects showed good adjustment, 11 showed fair, 12 were in the low-fair range, and 18 had an overall poor adjustment.

Further, the only therapeutic variables associated with outcome were those which described the severity of illness at the time of hospitalization. The greatest failure area among all subjects was social relationships. Only 21 percent could be described as having any type of warm, mutually gratifying relationships.

Additionally, Herrera et al. found that:

> The variables that were most strongly associated with long-range adjustment were leadership and "chumship" experiences before hospitalization--both key indicators of the ability to form relationships. All subjects with some such experience showed good or fair long-range adjustment, whereas every patient who had never been a leader or experienced chumships had a poor or low-fair long-term outcome (p. 773).

Studies of Concurrent Factors

Two studies collected information on treatment

within the facility and/or the characteristics of the patients during treatment.

Allerhand, Weber and Haug (1966) studied fifty boys who had been institutionalized at Bellefaire, a residential center for primarily Jewish adolescents. The data collection took place two years after discharge when the subjects were an average of eighteen years old. All of the subjects had been at Bellefaire for at least six months of treatment.

Allerhand et al. viewed "adaptability" as the crucial concept they were measuring; it was seen as central because

> a particular level of adaptability is the current integration of the individual's structural development with the resultant interaction between him and all the factors so far included in his life space (p. 140).

At follow-up, 71 percent of the boys were judged to be at least adequate in their overall adaption. When the milieu in which they were living was rated, 68 percent were judged to be in situations that were supportive.

However, the authors also reported that

> perhaps the most striking finding of the study is that none of the measurements of within Bellefaire performance at discharge, either in casework or in cottage and school roles, were useful in predicting post-discharge adaptability and adaption (p. 140).

A deficiency of this study was the lack of a control group which limited its ability to evaluate the success of Bellefaire's program.

Novotny and Burstein (1974) emphasized the significance of the post-discharge environment. The authors surveyed 94 delinquents who had been released from a juvenile corrective institution. Historical and background information, as well as test batteries and interviews, were used for the follow-up.

They found that of the 94 boys, 68 returned to high school, but 53 of these eventually dropped out. Fourteen boys graduated from high school. They found that school performance was associated with recidivism rate. Specifically, 50 percent of the boys who did not attend school had a subsequent felony conviction, but only 14 percent of the ones who graduated had such a conviction.

The authors attempted to analyze case history material of the 14 boys who did get through high school. They found that these boys did not differ significantly from drop-outs in regard to I.Q. or personality testing that had been done while they were in the corrective facility. However,

> most of the boys who graduated lived in a structured, supervised situation or had received supervision and encouragement from particular people in the environment while in school. Six of the 14 boys had finished public high school while living in a non-penal setting in cottages with other boys and houseparents. . . . Six of the other eight boys had help and attention from people or activities in the community. The people who were helpful, according to the boys' reports, included family members, foster parents, and social workers or probation officers (p. 55).

Of the boys who dropped out of school, more than half seemed to have received no outside support from people or programs. Expectations that these boys could succeed in adjustment without special aftercare arrangements were not realistic.

Studies of Both Antecedent and Concurrent Factors

Davids et al. (1968) conducted a follow-up study on a total of 37 children who were discharged between 1955 and 1964 from Bradley Hospital in Rhode Island. The authors utilized information about intake symptomology, case records during hospitalization, and a follow-up questionnaire filled out by parents designed to ascertain overall adjustment.

The final results focused on 27 patients, all males, about whom information was received. Of the 27, 10 had been diagnosed as childhood schizophrenics and 17 as passive aggressive personality disorders. The full scale mean I.Q. for passive aggressive disorders was 90, as compared to 75 for schizophrenics. Of the passive aggressives, 77 percent were discharged from the hospital with a favorable prognosis in comparison to 10 percent of schizophrenics. The two sub-groups did not differ on treatment variables such as kinds and amounts of drugs and psychotherapy they had received.

At follow-up, an average of five years after discharge, it was found that one-half of the subjects in both diagnostic categories were in institutions. Roughly one-third of the total sample had made a good adjustment.

The only significant difference that could be found in treatment variables between subjects who were institutionalized and subjects who were doing well was that more of the parents of successful cases had been seen in psychiatric casework (67 percent versus 17 percent).

The authors reported results from Eisenberg (1957) who studied 63 cases of infantile autism and Eaton and Menolascino (1967) who followed up 32 psychotic children who had been hospitalized. In both of these studies, no correlation was found between therapy, treatment variables, and eventual outcome.

Davids remarks that

> in several follow-up investigations, the best predictors of later adjustment were the chief complaint and presenting symptoms at the onset of treatment . . . our follow-up study fails to reveal any correlation between formal psychiatric treatment and clinical outcome (p. 475).

His conclusion is that the major factors which determine outcomes in residential treatment may have much more to do with the behaviors that patients bring with them to the treatment setting than with the actual therapeutic intervention.

Garber (1972) did a follow-up study of adolescent

boys and girls who were hospitalized between 1958 and 1968 at the Psychiatric and Psychosomatic Institute of Michael Reese Hospital in Chicago.

Four developmental tasks, derived from a psychoanalytic theory of adolescence were used to develop systematic ratings of post-hospital adjustment. Hospital records of 120 patients were examined in detail, and 71 of these were eventually interviewed face to face.

Results showed that 45 patients were functioning very well, 46 were functioning moderately, and 24 were functioning at a very low level.

In a latter stage of the research, chi square analysis was utilized to sort out the relationship of hospital experiences to overall adjustment. It was found that the two best predictors of current adjustment were lack of medication in the hospital and involvement with and interest in staff. In other words, patients who were healthier at admission also did better at follow-up, similar to what was previously reported by Davids et al. and Levy.

With regards to Garber's study, Durkin and Durkin (1975) have provided some criticism:

> While the study represents a major step, compared to previous outcome studies, in specifying the measures of functioning and hospital experience, its conclusions are limited by the lack of a control group. The true significance of the relationships is unclear because of the lenient 10 percent level of significance,

> especially since a large but unspecified number of chi square tests was conducted. The author, however, did not opt for such a rigorous research project; and given his limited goals he appears to have achieved them, namely, to have combined a clinical study and systematic data collection (p. 285).

Taylor and Alpert (1973) examined post-discharge adaption of children after residential treatment at Children's Village, a State-run facility in Connecticut.

Seventy-five children participated in the study. A version of the Roen-Burns Community Adaptation Scale (Roen & Burns, 1968) was utilized. Four hypotheses were tested:

1) The greater the degree of continuity of post-discharge environment, the greater the degree of the child's adaptation to the environment.
2) The greater the degree of the support in the post-discharge environment, the greater the degree of child's adaptation to the environment.
3) The greater the degree of pre-admission adaptation, the greater the degree of post-discharge adaptation.
4) The greater the degree of adaptation gained in the institution, the greater the degree of post-discharge adaptation.

Hypotheses one and two were supported while three and four were rejected, again similar to results reported

earlier. Adaptation while in treatment was found to be largely unrelated to post-discharge adaptation, with the exception of parent-child contacts during placement as well as staff contacts with family during placement. In terms of implications, the authors stressed the need for continuous family involvement prior to and during treatment and aftercare.

In a broad analysis of the results and methodology of treatment, Maluccio (1974) studied 215 children placed in residential treatment by the State of Rhode Island between 1964 and 1970. A total of 38 treatment centers, largely in the New England area, were represented by the 215 children.

Data were gathered in respect to: (1) characteristics of children and their families; (2) patterns of referral, placement, treatment, discharge, and aftercare; and (3) interrelations among key systems, especially referral agencies and treatment facilities. A large amount of data was compiled, the most significant of which will be summarized.

In regard to onset of illness in relation to onset of treatment, it was found that in over two-thirds of the cases, placement of the child in a residential facility occurred three or more years after his problems were first recognized by a community agency or professional person. This led the author to surmise that:

> There is considerable indication that parents and children in the years prior to the placement went from one community agency to another in a seemingly desperate search for solutions to their problems. The question must therefore be raised as to how many children could have been helped more effectively at home rather than through residential treatment if better previous services had been available in the community (p. 230).

With regard to outcome, 125 of the 215 children had been discharged as of the research. One-third of these had been discharged by the treating facility as successful, and one-third made no progress whatsoever, with the remainder somewhere in between.

It was possible for the study to evaluate actual progress of 76 children. Of these, 28 were substantially improved, 8 moderately improved, 17 minimally improved and 23 had shown no progress. The first two categories combined represent 48 percent of the population showing some worthwhile improvement.

Maluccio commented that most institutions had no organized procedure for gathering data on the child's adjustment following discharge.

His overall summary of the situation was that:

> the system of residential treatment seems characterized by numerous problems and limitations, particularly in respect to referral and admission patterns, discharge and aftercare services, participation of parents in treatment, and program effectiveness (p. 233).

Maluccio's study is commendable for surveying a large number of subjects in various treatment centers and

amassing a great deal of data. However, the study would have been more effective had it utilized as a control group a number of children who had not been placed in residential treatment but presented similar problems. Also lacking is data regarding those characteristics that might have differentiated children who improved from children who showed no progress.

Coche and Thomas (1975) evaluated the effectiveness of the Young People's Unit, a therapeutic community for adolescents at Friends Hospital in Philadelphia. This Unit holds 21 adolescent boys and girls, ages ranging from 14-21. A Sullivanian approach is utilized and the average length of stay is 50 days.

The authors used the Offer Self-image Questionnaire for adolescents (OSIQ, Offer, 1972). The OSIQ was given after admission, at discharge, and again one year after discharge. All in all, there were 137 subjects, 78 female and 59 male. Fifty-five were diagnosed as psychotic and 82 as non-psychotic. The scales of the OSIQ included: (1) impulse control; (2) emotional tone; (3) body and self-image; (4) social attitudes; (5) morals; (6) sexual attitudes; (7) family relationships; (8) external mastery; (9) vocational and educational goals; (10) psychopathology; and (11) superior adjustment.

Comparing the discharge to admission scores of the 79 patients who remained in the study, tests showed scales

five and six to be significant in the direction of improvement at the .01 level; all other scales were significant in the same direction at the .001 level, an impressive finding.

The conclusion of the authors was that the study

> demonstrates the powerful short-term beneficial effect of the hospitalization on the teen-agers' personal sense of adjustment and well being. The original question, whether the adolescent unit is helpful in improving the youngsters' adjustment, can be answered affirmatively. The mean OSIQ scores of patients do indeed improve dramatically and, as the follow-up OSIQ's of representative samples indicate, the improvement is maintained over time (p. 328).

One major weakness, however, of the study is that there is no control group. A second weakness would seem to be the lack of other measures to validate the results obtained by the OSIQ. The OSIQ has not frequently been utilized in the literature.

Studies Utilizing Control Groups

In two studies, a control group was used so that the effects of the treatment variable could be isolated. The combination of evaluation at the end of treatment with a follow-up study is a particularly powerful design.

Persons (1967) conducted a one-year follow-up study of the community adjustment of 41 delinquent boys, who, while incarcerated, had each participated in 40 group and 20 individual therapy interviews. He also utilized 41

matched control subjects who received no therapy.

In his initial study (1966) Persons had matched his subject and control groups on age, intelligence, race, socioeconomic background, type of offense, number of offenses, total time incarcerated, and nature of institutional adjustment. The mean age and I.Q. was 16.4 and 99.2 for the treated group and 16.3 and 97.6 for the control group. Each boy in the treated group had 80 hours and 60 sessions of group and individual therapy over a 20-week period. Five psychotherapists conducted the interviews, but boys always had the same group and individual therapist. The control group participated in the regular institutional milieu but received no therapy.

The results of the initial study showed that the therapy group made a superior adjustment as measured by psychological tests and various measures of overt behavior:

> The therapy boys showed better institutional adjustment, better interpersonal relationships, better performance in the institutional school, had fewer disciplinary reports, and received their institutional passes sooner than did boys in the control group (p. 138).

The follow-up study had a mean elapsed time of 9.5 months from the day of release to the date of data collection. The results showed that only 13 of 41 boys in the therapy group had been reinstitutionalized, whereas 25 of 41 boys in the control group were back in institutions. Twenty of 41 boys in the therapy group had violated parole, whereas 32 of the 41 control subjects had violated parole.

The $\underline{z}$ Tests between the differences in the proportions were significant at the .01 level.

Persons concluded on the basis of the above and other data that:

> The results of this study seem to indicate that psychotherapy can be an important factor in rehabilitation of delinquent youth. It should be particularly noted that only 5 of the 30 boys who were judged to be successfully treated subsequently became reinstitutionalized. However, from these results it should not be construed that psychotherapy is a rehabilitative panacea. For maximum results it seems that a boy needs to have a successful therapy experience, a reasonably adequate community replacement, and employment (p. 141).

Goldenberg (1971) conducted one of the more rigorous outcome studies and made use of an experimental and control group, as well as pre and post measures of both attitudes and behaviors.

The study was conducted at the Residential Youth Center which housed 20 adolescent boys who resided at the center and received group and individual therapy. For funding reasons, subjects in the study could not be randomly assigned. The experimental group consisted of the 25 most troubled youths, while the control group consisted of the next most troubled 25, who were not placed in a residential setting.

The methodology consisted of a structured interview analyzed in detail and measuring the dimensions of self concept, alienation, attitudes towards parents and authority, need for affiliation and need for achievement.

Differences in results between the experimental and control groups revealed that the experimental group became less alienated ($p < .01$), less authoritarian ($p < .05$), more trusting ($p < .10$), and more positive in views about the world ($p < .05$). Differences between control and experimental subjects prior to entering the center had been non-significant on all of the above scales.

Another behavioral measure used in the evaluation was the comparison of the number of days spent in jail. In the nine months prior to the opening of the RYC, the boys in the RYC spent 153 days in jail as compared with the control group's 140 days. In the nine months after the opening of the youth center, the RYC group spent 70 days in jail, a decrease of 54 percent, and the control group spent 258 days, an increase of 85 percent. A longer term post-discharge follow-up of the boys involved in this program has not yet been reported.

Related Studies

The second area of the literature to be reviewed consists of miscellaneous studies not categorized primarily as outcome or follow-up evaluations. These will be further broken down into correlational, descriptive or experimental studies.

Four of the studies in this section found significant treatment effects to be related to either group

psychotherapy (Tortorella, 1973; Truax, Wargo & Silber, 1966) or other forms of group treatment intervention such as role modeling or topic related discussions (Chandler, Greenspan & Barenboim, 1974; Sarason & Ganzer, 1973). The studies by Truax et al., Sarason et al., and Chandler et al. all utilized control groups and were among the best designed in the entire literature on residential treatment.

Polsky and Claster (1968) did a primarily correlational study utilizing a social systems approach and data they collected seem to have broad implications for future directions in treatment as well as research.

Rossman and Knesper (1976) did a descriptive analysis of several cases treated by means of behavior modification in a hospital setting. Cochrane's (1974) results showed that a correctional experience did not serve to positively influence the value systems of delinquents.

Each of the above studies as well as several more will be discussed more fully in the following sections.

Correlational Studies

Lefkowitz (1966) did a correlational study of the MMPI scores of 42 boys in treatment at the Berkshire farm for delinquents in New York.

One group of 21 boys was selected because of their good adjustment after six months in the program. The other group, also 21 boys, was selected on the basis of poor

adjustment. The mean age of all the participants was 14.5 and the mean I.Q. was 98. Adjustment was measured on the basis of the type of discharge that the subject received. The boys in the success group had received a regular discharge while the boys in the failure group had received a premature discharge due to acting out.

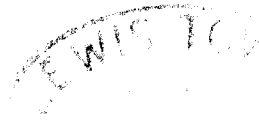
It was hypothesized that scales 6, 8, and 9, which as a group are intended to measure psychotocism, would be elevated for the boys in the failure group.

The hypothesis was partially confirmed. Only scale 9 (hypomania) was elevated at a significant level. Scales 6 (paranoia) and 8 (schizophrenia) differed in the hypothesized direction, but did not reach statistical significance. The authors felt that the results tended to support the hypothesis of greater psychopathology among the failure group, but cautioned that:

> The current findings are limited, however, because of the ex-post-facto nature of the research design. . . . Consequently other independent but unknown variables may have affected the outcome, albeit such common variables as age, social class and intelligence have been controlled (p. 913).

Polsky and Claster (1968) conducted an analysis of the social system of a residential treatment center.

The study was based on comparisons of three cottages at the Hollymeade center for Jewish delinquents in New York. An attempt was made to observe all aspects of cottage life systematically and comprehensively. The



final observation schedule consisted of 31 items and an observer's instruction manual which specified guidelines for coding and recording events. The observations focused on social roles of residents as well as treatment personnel at the institution.

There were four major roles that were attributed to child care workers. These four were: (1) nurturer and comforter; (2) counselor, guide and teacher; (3) mediator, integrator, and custodian; and (4) monitor and supervisor. Staff variability within the three cottages was examined by analyzing the content of staff-child interaction in terms of individual variations in different cottages.

This study has been praised as having important implications for future research in the field:

> Their emphasis on the residential treatment program as a social system will go a long way toward achieving their goal of formulating the empirical work on a sufficiently broad theoretical base so as to render it applicable with minimum variation to small group systems in diverse other settings as well (p. 306).

Kahn and McFarland (1973) evaluated 54 consecutive admissions to a delinquent treatment facility, largely in terms of academic skills. Of the 54, 47 finished the program and were included in the final data pool.

The authors utilized the Culture Fair Scale of Intelligence, the Step Reading Test, the Jesness Inventory, and the High School Personality Questionnaire. The mean age of the subjects in the study was 12.8 and racial com-

position was 46 percent Black and 2 percent Indian. Thirty-seven percent of the subjects were living in a broken home at the time of admission and 72 percent came from homes that had a history of a disruption of a significant impact during the boy's life.

Results showed that the Culture Fair I.Q. rose from a mean of 82.6 at admission to 88.0 at discharge. The authors felt that the most clear-cut measure of significant improvement came in the academic area, where the group as a whole improved an average of approximately ten percentile points on the reading test in the period of institutionalization. Improvement was also noted on five of the scales of the Jesness Inventory: social maladjustment, value orientation, alienation, manifest aggression and the asocial index.

Tortorella (1973) investigated the effects of milieu and individual therapy on WAIS I.Q. scores as well as personality changes in young girls between the ages of 16 and 19 who were referred by the courts of New York and placed in the Villa Loretto School, a residential setting. Seventy-one girls were tested on admission and 39 of them eventually completed the program.

Each girl was assigned to a social worker and a living group in which there were about 20 other girls. Subjects were seen in individual therapy once a week. Twenty of them also participated in group therapy once

a week. After confinement, an average of 12.6 months later, five subtests of the WAIS were readministered, as well as the MMPI and the Manifest Anxiety Scale.

Results showed a rise in the mean full scale I.Q. of 6.4 points for the girls who completed treatment. There was no post-testing done on the 32 girls who didn't stay in the program. When the MAS and MMPI were readministered at the completion of the program, there were significant changes in eight MMPI scales--bringing the scores closer to a normal profile.

Tortorella stated that:

> A major area of improvement seems to have been in the S's morale and feelings of usefulness and hopelessness as indicated by the significant change in the Depression scale. . . . The overall picture suggests a relief from excessive worry and an improvement in their ability to make use of their assets (p. 291).

It is unfortunate that a post test was not done at some point on the 32 girls who didn't stay in the program, because it could have added to the significance of the results. This is especially true considering that no control group was utilized in this study.

Cochrane (1974) attempted to objectively assess the value systems of young offenders in a correctional facility. The original subject pool was all new admissions to adolescent correctional institutions in the State of Michigan for one year. The age range was 13-17.

The boys were tested on admission with the

Rokeach Value Survey. They were also tested upon discharge but only 50 percent of the subjects tested at admission were available for retesting. According to Cochrane,

> . . . the data show that a training school experience has little impact on value systems and what impact there is may be interpreted as a retardation of the development of a mature and independent set of values (p. 344).

In other words, the training school experience was detrimental to the majority of the subjects. Cochrane pointed out that "as value systems are considered important factors in the etiology of delinquent behavior, it is evident that successful correctional programs should produce observable changes in value systems" (p. 344). This one did not.

Description of Treatment Programs

In a non-experimental but nonetheless interesting approach, Rossman and Knesper (1976) presented a design for a program to engage the resistent adolescent who cannot be treated ordinarily because of the severity of the acting-out behavior.

The procedures were utilized in an eighteen bed in-patient unit for adolescents at the University of Michigan Hospital. Subjects were reinforced for behavior that was considered to be socially appropriate. Disruptive behavior was not rewarded. No data was collected in the

program, but several treatment goals that appeared useful were stressed. These included the reduction of feelings of persecution and torment by the adolescent, the offering of meaningful identification models from among treatment staff, and the ability to adapt to constantly shifting adolescent needs.

The authors present two case histories with successful outcomes and generalize from these that:

> The implication of the case studies presented is that disruptive adolescents may be rapidly engaged within a therapeutic enviroment, if we utilize an approach which combines behavior modification techniques with a dynamic formulation of patient behaviors . . . this environment as described subsequently offers its own gratifications by allowing the adolescent to feel less anxious, to form positive attachments, and to find gratification through mastery and achievement (p. 706).

The study appears to be of value in suggesting direction for further research in the use of contractual systems with adolescents.

In a descriptive report, Scallon, Vitale, and Eschenauer (1976) discussed a program of behavior modification applied to classroom learning at a residential center.

The program was based on a token economy system geared to giving positive reinforcement, rewarding acceptable behavior and ignoring unacceptable behavior unless it was prolonged in which case a time-out room would be utilized.

One-fifth of a boy's allowance could be earned

daily if maximum checks were earned on a classroom behavior check list which broke the school day down into half-hour intervals.

It was found that the program could not be self-sufficient or contained within the classroom unless there was follow-up in the living situation from child care workers. The program had the tendency of bringing child care and school personnel closer together.

The program has now been in existence for four years and the authors reported that fifteen boys have successfully adjusted to public school settings after beginning on the behavior modification system. They also stated that

> . . . students in the school have stated sponteneously and when queried that the agency school atmosphere is much calmer and more controlled than in the years prior to implementation of the program (p. 568).

No control group or objective procedure was utilized to evaluate the program in a more systematic manner.

Experiments

Four of the studies reviewed qualify as experiments, since the investigator systematically varied one or more variables.

Levinson (1966) utilized a Q-Sort matching method to assign boys to cottages at the National Training School in Virginia. There were four possible assignment methods: (1) random; (2) Q-Sort matching; (3) staff selection of

boys; and (4) natural (boy picks boy) selection. The Q-Sort utilized statements from the Edwards Personal Preference Schedule (EPPS). Altogether sixty statements were selected and the sorting was designed to identify boys and staff members whose personality characteristics were similar. The hypothesis was that boys matched with staff on the basis of the Q-Sort would fare better in treatment than any of the other three groups.

Comparisons between the boys in all four of the different cottages were made after five months. Criterion measures for the comparisons consisted of (1) school grades; (2) vocational grades; (3) number of minor misconduct reports; (4) number of major misconduct reports; (5) total number of boys receiving misconduct reports (per cottage); (6) number of boys paroled; (7) number of boys transferred as failures; and (8) cottage adjustment grades.

Overall the results showed that the Q-Sort selected cottage residents showed the best level of performance. The mean rank achieved by Cottage "Q" on the eight criterion measures was significantly better than that obtained by the other three living units.

Levinson's conclusion was that:

> To the extent that these results are generalizable, the findings suggest the advisability of taking both the patient's adjustment level and the degree of counselor-patient similarity into consideration when patients are entering into therapeutic relationships (p. 364).

It is unfortunate that Levinson's study has never been replicated or evaluated further. Too often, children in residential placement are assigned randomly to a particular cottage or unit without consideration being given to their personality structure as a variable that needs to be considered in intake. If a single system or test could be devised to match staff with children, it would be an extremely valuable diagnostic tool.

Of course, Levinson's study was done at one school utilizing a primarily delinquent population so results are not necessarily generalizable to more psychiatrically disturbed populations. But the study appears to have employed adequate controls.

In one of the better designed experiments in the literature, Truax et al. (1966) tested the hypothesis that juvenile delinquent girls could be positively affected by group psychotherapy. The subjects consisted of 70 girls at the Kentucky Village for juvenile delinquent girls.

Forty girls were in the experimental group and 30 served as controls. It was expected that the experimental subjects would show superiority over the controls in terms of their ability to get out of and stay out of the institution during a one-year follow-up. It was also hypothesized that they would show improvement in the C scale, measuring juvenile delinquency, of the Minnesota Counseling Inventory (MCI). Also to be measured was self-concept

utilizing the Butler-Haigh Q-Sort technique (Butler & Haigh, 1954).

From the experimental group, four groups of ten girls each were formed. The therapists utilized in these groups were selected because they had ranked highest in a previous evaluation in accurate empathy and non-possessive warmth. All of the groups met for 24 sessions, twice a week for three months. Control subjects received no group therapy. As a check on the methodology, group sessions were recorded and samples analyzed to confirm the high accurate empathy and non-possessive warmth factors.

The results confirmed the hypotheses. An analysis was done on the percentage of time spent out of the institution, prior to the experiment and at a one-year follow-up. Experimental subjects had spent much more time out of the institution than control subjects; the difference in fact being signficant at the .001 level. In other words, subjects who received group therapy had adjusted considerably better.

The change on the MCI was significant at the .05 level, also in favor of improvement for the therapy subjects. On the C scale, girls receiving therapy showed "greater understanding of the need for social organization, more responsible behavior and less rebelliousness toward authority" (p. 270).

The authors' conclusion was that:

> Positive changes are produced beyond those observed in a control group . . . that such greater changes for the therapy group are neither transient placebo effects nor simply changes in test-taking behavior in an attempt to please the therapist is indicated by the therapy group's greater ability to get out of the institution and stay out during the one-year follow-up (p. 273).

This study is replicable, employed adequate controls, and dealt with one aspect of residential treatment in a very thorough manner. It is a conclusive indicator that group therapy can be an effective treatment tool. More importantly, it is a model for excellence in experimental design in a residential context.

Sarason and Ganzer (1973) did an experiment utilizing 192 male first offenders between the ages of 15.5 and 18 residing at the Cascadia Juvenile Reception-Diagnostic Center in Washington State.

The authors set up two treatment groups and a control group, each consisting of 64 subjects, a very adequate N. The hypothesis was that subjects in the group that was exposed to a series of role modeling sessions using adult models would become more socially adaptive as a result of these observational learning experiences. In the modeling sessions on particular subjects such as how to apply for a job, how to handle problem situations, etc. Groups consisted of 5 boys and 2 leaders. Each group met for 16 sessions that were each an hour long.

A second treatment condition consisted of discussion groups. The sequence and content was similar to the modeling groups, but there was no actual modeling, only group discussion. These groups also met for 16 hour-long sessions.

The control group was exposed to the regular institutional program but had no special group meetings.

Tests utilized included Sarason's Test Anxiety Scale (1962), the Pd scale of the MMPI, the Gough Impulsivity Scale (1957) and Rotter's Internalization-Externalization Scale (1966). In addition, cottage staff provided weekly behavior summaries and some self-report inventories were also used.

Post-measures were collected at the conclusion of the experiment and were also obtained at a follow-up three years later.

Results showed that the control group and both treatment groups had not differed significantly on any of the pre-measures. By the time of follow-up, subjects in both experimental conditions showed favorable changes in their attitudes, self-concepts, and their rated overt behavior. The authors suggest that:

> The modeling and structured discussion approaches had greater concurrent and long-term effects on adolescent delinquents than did the normal program of a high quality institution (p. 448).

The modeling condition subjects did better than discussion subjects in several areas.

Three years after their arrival at Cascadia, 43 of the 192 boys in the sample had become recidivists, but there were more recidivists in the control group (22) than in the modeling (12) or discussion (6) groups ($p < .06$). Also recall of content and purpose of the group was higher at follow-up for the modeling group (79 percent) than for the discussion group (38 percent). Both discussion and modeling subjects showed a greater ($p < .05$) shift toward internalization on the I-E scale than did the control subjects.

Chandler et al. (1974) evaluated 125 institutionalized emotionally disturbed children in terms of their role-taking and referential communication skills.

There were three major hypotheses. The first was that institutionalized emotionally disturbed children are characterized by chronic social adjustment problems and exhibit marked developmental delays in the acquisition of role-taking and referential communication skills. The second was that these developmental deficits could be partially remediated through programs of communication and role-taking. It was also felt that the above changes in developmental deficits would be accompanied by measurable improvements in social competence.

Of the 125 children originally selected, 48 were chosen on the basis of having especially low scores in the area of social competence. Role-taking ability and com-

petence were measured by cartoon sequences which children were asked to describe.

The 48 subjects were divided into three groups of 16. The first group was enrolled in an experimental training group utilizing drama and films as vehicles for providing remedial training in deficient role-taking skills.

Another 16 subjects participated in a referential communication training program which met two hours a week. These subjects participated in a series of communication exercises that place heavy demands on effective referential communications as well as the opportunity to identify and correct communication failures. The remaining 16 subjects served as an untreated control group.

Results showed large and statistically significant group change effects for both experimental conditions. All three hypotheses were confirmed.

The authors feel the study had broad implications from some future directions in residential treatment:

> Taken in combination, these findings lend additional weight to the initial orienting assumption that constructs and methods originally developed for the normative study of socio-cognitive development may be usefully transported into the study and possible amelioration of serious social and emotional disorders of childhood (p. 552).

A twelve-month follow-up showed a trend for improvements in both test measures to be associated with improvements in social adjustment as rated by institutional staff.

This study, along with Truax et al. (1966), appear to come the closest to meeting criteria for research with respect to randomization as well as controls.

Discussion

Perhaps the most succinct summarization of the current status of research in residential treatment is provided by the Joint Commission on Mental Health which stated that "few residential programs evaluate the outcome of their work in rigorously designed, well controlled, scientifically objective studies" (1970, p. 273). The overall conclusion of this survey is strongly supportive of this statement. Only a small proportion of the existing studies would appear to fulfill the requirements of sound experimental design.

Feinsilver and Gudnerson (1972) have established the following six criteria for meaningful research in the area of treatment of schizophrenics. The criteria they have established are presented here because it is felt that they are applicable to research in residential treatment as well:

1) a well-defined patient population;
2) matched control groups, in comparable milieus, including a group receiving drugs plus psychotherapy;
3) a well-defined relatively homogenous and non-idiosyncratic therapeutic approach;
4) a carefully designed and selected group of therapists;

5) measures of outcome evaluating behavior in a number of settings and taking into account the patient's own subjective experience of himself;
6) long-term treatment and follow-up (p. 22).

Actually only two or perhaps three of the studies reviewed here could really be said to come close to meeting the above criteria. Yet all of the criteria are obviously as relevant to research in residential treatment as they are to schizophrenia, with the possible exception of the drug plus psychotherapy contingency.

One of the major problems of these studies is the lack of a group of matched control subjects who do not receive the experimental treatment. Sometimes ethical objections are used to justify this. Some researchers would say it is unethical to deprive some children of a potentially helpful treatment by placing them in an untreated control group. This argument seems to be putting the cart before the horse. Until the treatment has been proven to be successful, there is no reason to assume that it is worthwhile. Further, the issue of presenting uncontrolled research which becomes accepted into the literature also seems to be an ethical one.

Another problem is the lack of good description of the psychopathology and presenting problems of the child involved in the research. Although this survey purports to consider studies of so-called "emotionally disturbed" children, there is no assurance that the subjects in the various studies really have all that much in common.

Therefore one cannot really be sure if one is reviewing a body of literature pertaining to a particular form of illness, or several smaller groupings of studies pertaining to various illnesses.

A third problem in respect to these studies is the lack of uniform measuring instruments with established clinical validity. Each researcher is utilizing his own special tool but there is often no way to know if all of these tools are really measuring the same thing or what they are indeed measuring. A great need in this field would be the establishment of some instruments that are generally recognized in the field as being useful and valid. These instruments could then serve as validating measures for almost any experiment where additional confirmation was needed. It seems that a useful literature cannot be built unless there is a field-related methodology on which to build.

Lander and Schulman (1960) expressed the view that the evaluation of the effectiveness of a milieu is more difficult than the evaluation of the effectiveness of individual treatment because of the greater number of variables in the milieu. In addition, researchers, though well intentioned may not have a sound knowledge of group dynamics and influences of the institutional system on the individual. This may explain why many evaluations of residential programs have been descriptive and impressionistic; really

prerequisites to forming some hypotheses and conducting experiments.

Herstein (1975) asserts that there is a dichotomy in the field between the clinical point of view which most closely seeks to identify internally oriented change in the child, and the community point of view which seeks externally determined criteria. For community standards to be met, it may only be necessary for acting out or destructive behavior to cease, regardless of whether meaningful dynamic changes have been made that will sustain the behavioral change. It is Herstein's recommendation that researchers not become so completely swayed by budgetary accountability as to abandon the clinical point of view which he sees as more truly worthwhile.

Cohen (1969) has delineated the relationship of sample size to statistical power and feels that studies of residential treatment programs with fewer then ten children would preclude the finding of meaningful significant differences. However, as most residential centers seem to have populations of forty or more, it would not seem likely that this problem would hamper most researchers.

Maluccio and Marlow (1972) have commented that the definition of "successful outcome" in residential treatment is very loose. Matushima (1965) has elaborated:

> Current intake criteria . . . are not precise enough to permit measured comparison between settings. Also

> treatment "success" is defined in various ways. How much change, intrapsychic, in relation to environment, or both must there be for a case to be considered successfully treated? (p. 277).

The question that Matushima asks has not really been answered in a standardized way; therefore each individual experimenter seems to come up with his own answer to the question. Valid instruments with norms provided for healthy as well as disturbed children would be helpful in dealing with this issue.

Some authors (Etzioni, 1960; Schulberg & Baker, 1968) have focused specifically on follow-up studies and the difficulties inherent in this type of research. The major point is that there is too much delay in providing relevant feedback back to the program about its effectiveness. Also important is the problem of weighing the contribution of various contributing variables in the context of a whole social system, as well as the aftercare system.

Dubois (1973) adds that the follow-up studies need to be longitudinal rather than cross-sectional, as is usually the case. Cross-sectional studies

> . . . indicate the child's emotional status at a particular point in time only. This is not appropriate for studying the outcome of emotional disturbance--there is too much fluctuation in the natural history and course of the illness. Longitudinal follow-up is required to provide an adequate description of the child's post-discharge course (p. 3).

Maluccio and Marlow also recommend that further work is needed to identify specific factors in the post-institu-

tional environment that contribute to successful outcome. Follow-up studies cannot be done in a vacuum; they must consider the effect of the aftercare plan on the treatment.

It is good that residential centers will need to be able to justify their results in order to receive more and better funding. However, unless meaningful criteria and methodology for evaluation also comes with accountability, its effect will be spurious at best. This review offers little promise that the situation in regard to research is improving significantly at this time. All that is clear is that dramatic improvement is needed as soon as possible.

III. THE PROCEDURE

Setting

The study was conducted at Lawrence Hall School, a residential treatment center for emotionally disturbed boys on the North Side of Chicago. The 48 boys in the campus program at Lawrence Hall are generally wards of the States of Illinois or Indiana. Boys referred by the State to Lawrence Hall have generally been exposed to previous treatment interventions such as outpatient therapy, special education, and foster care. The cost of care at Lawrence Hall is expensive ($55.00 per diem) so that the State is likely to have tried a number of previous approaches before acknowledging the need for a residential setting.

Diagnostically, the boys most often but not universally fall into the category of behavioral or character disorders. Most of them have a history of school and community acting out that is reactive to the family situation. A very small percentage of the boys are psychotic or borderline, and occasionally there are boys who have more neurotic symptoms such as phobias or compulsive behavior. In some cases, the severe acting out of a character disordered boy has been associated with borderline

reality testing and ego controls. In such cases, medication may be used.

It was felt that results of this study would be generalizable for other populations of institutionalized adolescents whose primary symptoms are in the area of behavioral and characterological problems. These boys are most often pre-delinquent rather than severely delinquent so results are more generalizable for residential treatment centers that treat acting out adolescents as a last step prior to commitment to a juvenile correctional facility.

Boys under thirteen years old were not included in the study to comply with recommended age norms for the Devereaux Scale.

Subjects

The experimental and control groups both consisted of 21 boys between the ages of 13 and 18 who were given the Devereaux Adolescent Behavior Rating Scale a total of three different times. Three cottages of 7-8 boys were selected to receive the experimental condition. The control group consisted of appropriate aged boys from four different cottages.

All boys in any one cottage were exposed to the same intervention or lack thereof in a randomized group design, rather than cottages being split in terms of treatment versus no treatment.

Since random assignment of individuals to treatment conditions was impractical, matching was done after the fact, but on the basis of means obtained in the pre-testing on the five selected factors of the Devereux Scale: unethical behavior, defiant-resistive, hyperactivity, poor emotional control and inability to delay gratification. After pre-test means for cottages were obtained, t tests for each of the five variables were conducted to be certain the experimental and control group populations did not differ significantly. Results for the pre-test and resulting t tests are given in Table 1.

None of the t test levels approached significance, indicating that the experimental and control subjects did not differ as a population on any of the five Devereux variable.

The experimental group received the behavior modification program to be described in the Section on Method. The control group received no behavior modification interventions during the three-month intervention period. Both groups were tested after three months on the same five factors of the Devereux test, and a follow-up testing was done sixty days after the conclusion of the experiment. In addition a questionnaire was administered to staff members participating in the experimental group program. The results of this questionnaire were also evaluated.

Table 1

Cottage Pre-Test Means, Standard Deviations, and t Test Results for Experimental Versus Control Group Populations

	N per Cottage	Variable									
		Unethical Behavior		Defiant Resistive		Hyper-Active		Poor Control		Inability to Delay	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Experimental Group Cottages											
Graves	8	11.25	3.69	13.88	4.70	21.25	5.80	17.50	4.69	26.75	6.39
Drake	6	6.17	1.94	8.83	4.62	14.17	5.49	11.83	3.87	18.00	4.20
Ewen	7	14.29	4.89	14.29	1.60	22.57	5.00	20.00	2.94	26.86	5.93
Experimental Totals	21	10.57	5.00	12.57	4.44	19.62	6.31	16.71	5.02	24.29	6.77
Control Group Cottages											
Randall	4	13.25	5.74	13.50	3.87	19.00	5.72	19.00	4.08	25.00	6.88
Avers	5	11.60	5.73	9.80	3.83	19.20	3.19	14.40	3.51	17.40	3.29
Byron	6	15.33	5.05	16.83	1.47	22.17	3.87	20.33	3.33	30.83	5.91
Hoover	6	12.17	3.76	12.50	3.21	20.33	5.01	16.67	3.44	23.17	6.24
Control Totals	21	13.05	4.78	13.29	3.91	20.33	4.32	17.62	4.02	24.33	7.28
T-Test Scores for Experimental Versus Control Totals	42	1.62		0.54		0.42		0.63		0.02	

Instrument

The Devereux Adolescent Behavior Rating Scale (Spivak, Hames, and Spotts, 1967) is described as "a means whereby an individual who has intimate living contact with the youngster can reliably describe and communicate to others the wide range of overt symptomatic behaviors" (p. 3).

The Scale profiles fifteen problem behavior factors including unethical behavior, defiant-resistive, domineering-sadistic, heterosexual interest, hyperactive-expansive, poor emotional control, need approval and dependency, emotional distance, physical inferiority-timidity, schizoid withdrawal, bizarre speech and cognition, bizarre action, inability to delay, paranoid thinking, and anxious self-blame.

Under each of the above categories are several different items which the rater rates the subject on by using a numbered scale. A final score is recorded for each cluster of items or "factor" and an overall profile on all fifteen factors is also obtained. Appendix A provides a sample Devereux test that was administered to a Lawrence Hall resident, as well as an analysis of the resulting profile.

The period of observation of the Devereux test is the two weeks immediately prior to rating. It is suggested that the rater be someone who is intimately familiar

with the child's behavior on a day-to-day basis. The implication of this is that the child care worker is more qualified to do the rating than is a therapist or case-worker. Training should be provided for raters in using the Scale and this will be explained in the Section called Method.

For the purpose of this experiment, five of the factors on the Devereux Scale were selected as being particularly demonstrative of the types of problem behaviors that Lawrence Hall residents frequently display. These five are unethical, defiant-resistive, hyperactive-expansive, poor emotional control, and inability to delay. Since the majority of Lawrence Hall residents are diagnosed in the range of behavior disorders and impulse control problems, it was felt that these five factors reflect behaviors where there is a great need for therapeutic interventions.

The reason for data analysis being restricted to five of the sub-scales of the Devereux centers around the practical issue of establishing a behavior modification program that is directed towards target behaviors. Five such target behaviors resulted in the construction of a Point Reward Score Sheet (see Appendix B) of twenty items, and it was felt that this was complex enough, since this item must be administered on a daily basis by child care

workers. To have included more behaviors in the basic experiment would only make it cumbersome.

Normative data on the Devereux Scale is provided for the following diagnostic categories: passive-aggressive personality, childhood schizophrenia, retarded, adjustment reaction, chronic brain syndrome, anti-social personality, and neurotic. A total of 407 institutionalized disturbed adolescents were used in obtaining the normative data for the groups represented above. Data were also obtained for normal children living in foster homes and normal children living in their own homes.

The authors provide data on test-retest reliability for 89 subjects who were tested 7-10 days apart. The correlation obtained was .82 which the authors considered to be very good considering that some of the rated behaviors may actually change over a period of time.

Interrater reliability was established by having parents rate children living at home. Two parents would rate the same child at the same time. The coefficient of agreement was .90. Different staff members also rated 89 institutionalized adolescents and the coefficient of agreement was .81.

It was desirable in this experiment for the rater who did the initial rating of a resident to also do the first reevaluation after three months and the follow-up after sixty days.

Method

The study is basically a pre-post model with pre-measures being taken prior to assignment to experimental and control groups. Table 2 summarizes the design for the experiment. Complete random assignment of individual subjects to control and experimental conditions is not possible. Therefore, the study is technically an example of quasi-experimental randomized group design (Campbell and Stanley, 1970).

The procedure for the experiment was as follows. The initial subject pool consisting of 44 boys was given the Devereux Scale. Rating was done by trained child care workers who were employed full time in the unit where the rated subject resided. A total of ten different raters were utilized in the experiment. Five of the seven cottages were rated throughout the pre and post testing by the same rater which is the most ideal condition. In the sixth cottage, the initial rater resigned during the experiment and it was necessary to use a second rater. In the seventh cottage rating was split between two raters. Raters were aware whether or not their particular unit was in the experimental or control group. However, they were not specifically aware of how the experimental treatment was tied to performance on the Devereux Scale. Therefore, it is felt that their ratings were likely to be unbiased.

Table 2

Schematic Outline of the Design for the Experiment

Pre-Test	Intervention	Post-Test	Follow-Up
Experimental group: Given Devereux Test S's matched on five scales to control group	Intensive three-month Behavior Modification program	Devereux Test-Same 5 scales	Devereux Test again after 60 days
Control group: S's matched with Experimental Group on 5 scales of Devereux Test	No intervention	Devereux Test-Same 5 scales	Devereux Test again after 60 days

After the initial rating, profiles were drawn up and special attention paid to scores obtained on behavioral factors one, two, five, and six and rational cluster one of the Devereux Scale.

Three cottages were assigned to the experimental condition and four to the control condition. The assignment was random.

That there were 24 boys in each group allowed for some attrition during the three month experimental period. This attrition was due to discharge from Lawrence Hall, either planned or unplanned. Also, not all boys participated for an equal number of weeks in the experimental condition. This was due to occasional passes or runaways. This effected only a few of the subjects, however, and was generally for no more than one to two weeks out of the entire experiment. Table 17 on page 88 indicates the specific number of weeks that each subject actually participated in the experiment.

For boys assigned to the experimental condition, a behavior modification program was implemented, utilizing the "Point Reward Score Sheet" (see Appendix B). The Point Reward Score Sheet was devised specifically by the experimenter for use in this experiment. It consists of 20 items that are descriptive of adolescent behavior. It was constructed in order to shape behavior towards improvement on the five key behavior factors on the Devereux

Scale. In other words, boys could receive points leading to rewards for improving their behavior in five specific areas: ethics, defiance and resistance, hyperactivity, emotional control and ability to delay gratification.

On the Score Sheet, the first four items are correlated with the ethical behavior factor of the Devereux Scale. Items 5 through 8 deal with "defiant-resistive behavior," items 9 through 12 with hyperactivity, items 13 through 16 with emotional control and items 17 through 20 with delay of gratification.

The crux of the experiment was whether or not reinforcement of these same items consistently over a period of three months would lead to a measurable improvement in ratings obtained on the corresponding factors of the Devereux Scale.

The Point Reward Score Sheet was designed to be as simple as possible so that it could be administered on a daily basis by child care workers and readily understood by the subjects.

It seemed important that the Score Sheet be used uniformly by all staff members participating in the experiment: therefore, time was spent during staff meetings discussing scoring criteria for the various items. The overall procedure for training raters is explained below.

Child care workers reviewed the meaning of the items on the Point Reward Score Sheet with subjects on a

frequent basis to assist in shaping their behavior. It was suggested to them that they try to do this on a daily basis, and after the subjects are more familiar with the Score Sheet, somewhat less frequently. Each boy was told that his score would be computed at the end of each day at approximately 9 P.M. If he had earned it, he was to be given a reward listed under the "menu of reinforcement" (see Table 3).

Subjects were told that if they earned 650 points in a week, they would be eligible for a Certificate of Good Behavior (see Appendix D), as well as the choice between a trip to a restaurant or a trip to a dime store valued at $1.50.

After 12 weeks, subjects who earned it received the long term end reward of a special trip to dinner and a basketball game. Subjects were reminded throughout the entire experiment of the possibility of earning this desirable end result.

Praise was an important component of the reinforcement system. As much as possible, staff members participating in the experiment were encouraged to provide uniform positive responses for the achievement of certain target behaviors as reflected in the point totals.

At the end of the elapsed 12 weeks of intervention, both control and experimental groups were again given the Devereux Test. The results for the specified

Table 3

Menu of Reinforcement for the Experimental Group

Daily Rewards:	Each day with 90 or more points, subject may select from: praise + 1 candy bar or 1 comic book or 1 of assorted 20-cent items such as pocket combs, life-savers, baseball cards, chewing gum, etc.
Weekly Rewards:	650 points or more in one week, subject received: Certificate of Good Behavior + praise + choice of: 1) trip to hot dog or hamburger stand--able to spend up to $1.50 for food. 2) trip to dime store or toy store--able to buy item of choice up to value of $1.50.
Final Reward:	Ten Certificates of Good Behavior in a 12-week period entitles subject to a trip (in a group of all subjects who have earned it) to a professional basketball game or equivalent sports event.

scales were analyzed. The hypothesis was that the group which had received specific rewards for certain behaviors that were correlated to items on the Devereux Test would show more improvement on those items. This would indicate that the behavior modification system was effective in shaping certain desirable target behaviors.

After 60 days, the Devereux Test was again given to both control and experimental groups to determine any changes from the intitial pre and post test results.

The training procedure for the raters using the Devereux Scale was as follows. All of the raters selected were called together for a meeting. At the meeting, the rating guide on the front page of the Devereux Scale (Appendix A) was reviewed. The guide is fairly concrete and provides very unambiguous directions. However, where necessary, sample ratings for various subjects were discussed by the raters. Training sessions took about one hour.

Prior to the second and third Devereux ratings, short review sessions were held with all raters in order to cover the basic procedures and cover any questions that may have arisen.

Raters remained naive as to the design of the experiment and their exact role in it. They were told that they would receive a written explanation of the

experiment at the conclusion of their participation in it.

A completely separate type of training was necessary for the utilization of the Point Reward Score Sheet (PRSS) (Appendix B). In this case there were six raters involved, two from each of the three experimental group cottages. It was necessary to have at least two raters from each cottage because the program is a seven day program--but each staff member has two days off during the week; therefore, a second staff member was necessary to rate the boy on the days the first staff member is off. There was no need for any raters from the control group cottages.

Raters on the PRSS were asked to do their evaluation as late in the day as possible. They were asked to rate the subjects according to their own best judgment, although they may rely on reports of behavior from other staff members. It was also emphasized that scoring should be done on an all or none basis to keep the daily administration of the program as simple as possible.

There is certainly some subjectivity involved in using this type of a rating scale, but raters were asked to adhere as closely as possible to rating on the pure behavior of the boys involved. In the training sessions, key terms from the PRSS were defined in terms of their behavioral correlates. Some of the more ambiguous terms required further discussion. Obviously,

some of the items on the PRSS are generally easier to rate than others.

The raters involved in the study were experienced child care workers who were familiar with the children involved. Although part of their role is to be an advocate and supporter of the boys, these staff members are capable of limit setting and objectivity with respect to the child's behavior. They are often put in a position of having to be firm with a boy or to deny a request under pressure. It seems unlikely that they would be biased in terms of helping kids to achieve on this scale, other than by encouraging them to do their best to obtain the desired rewards. It was explained to these raters that objectivity on their part was necessary for the successful implementation of this behavior modification program. If they were to shade the results in any way, the children involved would ultimately not benefit from the program.

The experimenter attended staff meetings on the experimental units approximately every other week during the twelve weeks of the program. This enabled him to discuss with staff any questions that arose regarding the utilization of the PRSS.

Statistically, data were analyzed using a simple one-way analysis of variance for cottage groups across each of the main Devereux factors. This was to test the

hypothesis that the behavior modification program produced improvement in the subject's behavior on the various criteria.

IV. RESULTS

The results of the experiment were as follows: A simple one-way analysis of variance was performed across all five Devereux variables with the experimental versus control group as the main effect. For the three month time interval from pre-test to post-test, raw scores and group means are shown in Tables 4, 5, 6, and 7.

As the five Devereux variables were non-linear, *z* scores were first computed and utilized in the *F* computation. Table 8 presents the grand means and significance levels for the pre to post-test time interval. A Bartlett test performed for all five of these variables indicated no significant heterogenity of variance.

As indicated in Table 8, none of the *F* levels obtained approach significance. This means that taken as whole samples, the experimental and control groups did not differ significantly. The basic hypothesis that the experimental group would improve more on the Devereux variables than the control was therefore not supported over this time interval. Statistically significant results, however, were obtained when a follow-up test was given sixty days later and these results are discussed later in this section.

A second analysis of variance was run for all seven

Table 4

Experimental Group Pre-Test Raw Scores and Cottage Group Means Across All Five Dependent Variables

Subject	Cottage Group	Variable				
		Unethical Behavior	Defiant Behavior	Hyper-activity	Poor Control	In-ability to Delay
1	1	6	5	21	9	17
2	1	7	17	10	17	25
3	1	12	11	23	16	23
4	1	18	20	30	24	33
5	1	13	14	19	18	29
6	1	11	12	20	18	28
7	1	11	18	26	23	37
8	1	12	14	21	15	22
9	2	6	7	16	14	21
10	2	5	17	15	16	23
11	2	5	6	14	13	14
12	2	4	8	7	7	13
13	2	8	4	23	7	21
14	2	9	11	10	14	16
15	3	9	12	25	21	30
16	3	19	16	19	22	29
17	3	10	13	32	16	27
18	3	19	16	20	18	18
19	3	10	15	21	25	35
20	3	20	15	17	19	20
21	3	13	13	24	19	29
Group 1 Mean		11.25	13.88	21.25	17.50	26.75
Group 2 Mean		6.17	8.83	14.17	11.83	18.00
Group 3 Mean		14.29	14.29	22.57	20.00	26.86

Table 5

Experimental Group Post-Test Raw Scores and Cottage Group Means Across All Five Dependent Variables

Subject	Cottage Group	Variable				
		Unethical Behavior	Defiant Behavior	Hyper-activity	Poor Control	In-ability to Delay
1	1	8	10	22	15	22
2	1	8	16	11	15	21
3	1	14	16	23	14	26
4	1	17	17	24	19	31
5	1	18	18	21	16	29
6	1	18	15	20	20	31
7	1	7	13	29	18	29
8	1	17	17	26	17	26
9	2	5	16	15	17	24
10	2	4	9	9	11	19
11	2	8	14	11	13	22
12	2	6	8	7	7	14
13	2	12	6	11	5	17
14	2	4	5	8	8	11
15	3	10	14	18	22	28
16	3	18	14	18	22	28
17	3	7	11	25	17	28
18	3	13	12	15	13	14
19	3	8	13	19	22	32
20	3	14	10	15	13	14
21	3	14	14	26	17	29
Group 1 Mean		13.38	15.25	22.00	16.75	26.88
Group 2 Mean		6.50	9.67	10.17	10.17	17.83
Group 3 Mean		12.00	12.57	19.14	18.00	24.71

Table 7

Control Group Post-Test Raw Scores and Cottage Group Means Across All Five Dependent Variables

Subject	Cottage Group	Variable				
		Unethical Behavior	Defiant Behavior	Hyper-activity	Poor Control	In-ability to Delay
22	4	12	14	18	17	17
23	4	21	7	17	16	25
24	4	7	9	15	18	25
25	4	16	11	17	14	22
26	5	6	10	21	18	22
27	5	9	11	15	15	19
28	5	7	9	18	10	12
29	5	15	8	23	16	21
30	5	23	18	21	20	27
31	6	12	14	24	21	24
32	6	18	17	17	19	33
33	6	21	18	20	17	20
34	6	18	19	24	24	29
35	6	7	15	23	24	42
36	6	8	11	12	13	18
37	7	16	14	22	21	30
38	7	11	11	27	16	23
39	7	7	10	19	14	21
40	7	15	19	23	23	34
41	7	7	9	19	9	14
42	7	11	16	13	19	22
Group 4 Mean		14.06	10.25	16.75	16.25	22.25
Group 5 Mean		12.00	11.20	19.60	15.80	20.20
Group 6 Mean		14.00	15.67	20.00	19.67	27.67
Group 7 Mean		11.17	13.17	20.50	17.00	24.00

Table 8

Grand Means and Significance Levels Pre to Post-Test for Experimental Versus Control Groups Across Each of Five Main Dependent Variables Separately

	Unethical Behavior	Defiant Behavior	Hyper-Activity	Poor Control	Inability to Delay
Experimental Group					
Grand Pre-Mean	10.57	12.57	19.26	16.71	24.29
Grand Post-Mean	10.95	12.76	17.67	15.29	23.57
Control Group					
Grand Pre-Mean	13.05	13.29	20.33	17.62	24.33
Grand Post-Mean	12.71	12.86	19.42	17.33	23.81
F	.37	.21	.80	1.35	.02

Note. Df = 1, 40

cottages, pre to post test considering cottages as a main effect. Results of this ANOV are presented in Table 9.

In this case, there is a very weak tendency ($p < .10$) for variable three to be significantly different across cottages. The implications of this are discussed below. All other variables do not differ across cottages.

The Bartlett test indicated significant heterogenity of variance for variables two ($p < .03$) and three ($p < .03$).

A third analysis of variance was conducted for groups within the experimental population only as the main effect. The results of this are shown in Table 10.

Because two of these variables show a significance level approaching the $p < .05$ level, a breakdown was computed as to the direction and magnitude of z score means for each of the cottages on each of the five variables. Table 11 summarizes these means.

The Bartlett Test indicated no significant heterogenity of variance for any of the above five variables.

An inspection of the means in Table 11 shows that Group Three, Ewen, which was the youngest cottage shows a positive change across all five variables. Group Two showed a positive change for two of five variables and Group One showed a negative change in all five measures.

Table 9

Analysis of Variance Pre to Post Test Across Seven Cottages for Each of Five Main Dependent Variables Separately

Variable	F	Significance Level
Unethical Behavior	1.40	(NS)
Defiant Behavior	1.00	(NS)
Hyperactivity	1.99	.09
Poor Control	1.00	(NS)
Inability to Delay	.92	(NS)

Note. Df = 6, 35

Table 10

Analysis of Variance Pre to Post Test, Experimental Group Cottages Only, for Each of Five Main Dependent Variables Separately

Variable	F	Significance Level
Unethical Behavior	3.44	.05
Defiant Behavior	.94	(NS)
Hyperactivity	3.40	.06
Poor Control	.28	(NS)
Inability to Delay	.60	(NS)

Note. Df = 2, 18

Table 11

Direction and Magnitude of z Score Means for Experimental Group Cottages Across All Five Dependent Variables

Variable	Cottage Group	Mean
Unethical Behavior	1: Graves	-.44
	2: Drake	-.14
	3: Ewen	+.43
Defiant Behavior	1: Graves	-.43
	2: Drake	-.14
	3: Ewen	+.39
Hyperactivity	1: Graves	-.39
	2: Drake	+.46
	3: Ewen	+.32
Poor Control	1: Graves	-.02
	2: Drake	+.18
	3: Ewen	+.25
Inability to Delay	1: Graves	-.12
	2: Drake	-.03
	3: Ewen	+.21

In variable one, unethical behavior, where the .05 level of significance is closely approached, subjects in group three showed the greatest changes in the positive direction. In variable three, hyperactivity, where the .05 level is also approached group two subjects and group three subjects both had changes in the direction of improvement.

These results imply that when the experimental group population is broken down into cottages for the pre to post-test interval, there is a trend indicating that Ewen Cottage subjects differed significantly from the other two groups in their scores and that this change is in the direction of improvement on all five variables. The change approaches the .05 level of significance for the variables of unethical behavior and hyperactivity. There is a slighter trend across all five variables for Group Three subjects to show some change in a positive direction. Group Two subjects, also showed a change approaching significance on the variable of hyperactivity.

Since cottages as groups are quite small, with an N ranging from six to eight subjects per cottage, it would be safe to say the replication with larger groups would add more significance to these results. There is no doubt, however, that grounds exist which would justify further research with groups of this sort.

Results of the Follow-Up Rating

A follow-up rating on the five Devereux factors was conducted sixty days after the completion of the experiment. Tables 12 and 13 present raw scores and means for this rating.

A simple one way analysis of variance was conducted comparing control and experimental group populations across all five variables from pre-test to follow-up test. Table 14 summarizes these results.

The Bartlett test performed for these variables indicated no significant heterogenity of variance.

These results indicate that improvement in target behaviors had taken place in the sixty days after the experimental condition ceased, and that a significance level of $p < .01$ differentiated the experimental and control populations for variables of poor emotional control and inability to delay gratification.

The improvement was further confirmed by the post-test to follow-up analysis of variance, the results of which appear in Table 15. It is clear from this test result that the heaviest change in the test scores did in fact occur between the post and follow-up test as the significance level for variable four is $p < .05$ and for variable five is $p < .01$.

Implications of this delayed effect are discussed in Chapter V.

Table 12

Experimental Group Follow-Up Test Raw Scores and Cottage Group Means Across All Five Dependent Variables

Subject	Cottage Group	Variable: Unethical Behavior	Defiant Behavior	Hyper-activity	Poor Control	In-ability to Delay
1	1	7	9	21	14	19
2	1	7	13	10	14	21
3	1	10	14	19	12	22
4	1	14	14	19	18	26
5	1	16	15	17	14	24
6	1	15	14	16	17	23
7	1	9	15	21	14	24
8	1	22	19	26	18	33
9	2	11	20	15	17	30
10	2	5	18	14	19	32
11	2	6	5	11	6	14
12	2	8	12	7	7	10
13	2	9	5	16	7	16
14	2	6	12	26	13	17
15	3	9	9	15	15	18
16	3	14	12	17	16	21
17	3	7	9	22	12	22
18	3	16	15	20	16	19
19	3	9	13	16	22	24
20	3	12	11	19	12	16
21	3	15	12	25	12	26
Group 1 Mean		12.50	14.13	18.63	15.12	24.00
Group 2 Mean		7.50	12.00	14.83	11.50	19.83
Group 3 Mean		11.71	11.58	19.14	15.00	20.86

Table 13

Control Group Follow-Up Test Raw Scores and Cottage Group Means Across All Five Dependent Variables

Subject	Cottage Group	Variable: Unethical Behavior	Defiant Behavior	Hyper-activity	Poor Control	In-ability to Delay
22	4	11	9	15	14	21
23	4	22	11	20	17	28
24	4	9	11	17	17	20
25	4	12	9	21	15	17
26	5	7	12	19	17	21
27	5	8	10	16	16	23
28	5	7	9	16	10	13
29	5	12	12	22	14	23
30	5	23	18	18	20	27
31	6	17	17	20	22	30
32	6	19	19	18	17	30
33	6	22	20	22	20	32
34	6	18	17	27	24	32
35	6	10	17	31	22	38
36	6	19	16	17	18	29
37	7	23	20	28	25	40
38	7	18	20	25	25	38
39	7	15	16	19	23	33
40	7	16	20	24	25	39
41	7	8	8	16	11	15
42	7	17	20	12	20	31
Group 4 Mean		13.50	10.00	18.25	15.75	21.50
Group 5 Mean		11.40	12.20	18.20	15.40	21.20
Group 6 Mean		17.50	17.67	22.50	20.50	31.83
Group 7 Mean		16.16	17.33	20.67	21.50	32.67

Table 14

Grand Means and Significance Levels Pre-Test to Follow-Up for Experimental Versus Control Groups Across Each of Main Dependent Variables Separately

	Unethical Behavior	Defiant Behavior	Hyper-activity	Poor Control	Inability to Delay
Experimental Group					
Pre Test Grand Mean	10.57	12.57	19.26	16.71	24.29
Follow-Up Test Grand Mean	10.81	12.67	17.71	14.05	21.76
Control Group					
Pre Test Grand Mean	13.05	13.24	20.33	17.62	24.33
Follow-Up Test Grand Mean	14.90	14.81	20.14	18.67	27.57
F	1.56	1.06	1.76	8.15*	10.99*

*$p < .01$

Note. *Df* = 1, 40

Table 15

Grand Means and Significance Levels Post-Test to Follow-Up for Experimental Versus Control Groups Across Each of Five Main Dependent Variables Separately

	Unethical Behavior	Defiant Behavior	Hyper-activity	Poor Control	Inability to Delay
Experimental Group					
Post Test Grand Mean	10.95	12.76	17.67	15.29	22.57
Follow-Up Test Grand Mean	10.81	12.67	17.71	14.05	21.76
Control Group					
Post Test Grand Mean	12.71	12.86	19.42	17.33	23.81
Follow-Up Test Grand Mean	14.90	14.81	20.14	18.67	27.57
F	3.97	2.80	.47	4.52*	10.99**

*$p < .05$

**$p < .01$

Note. $Df = 1, 40$

Inspections of $\bar{z}$ score means contributing to the above significance levels revealed the cottage by cottage source of the experimental group improvement. These means are presented in Table 16.

Inspection of these means reveals that the significant source of the improvement in the experimental group occurred in Graves Cottage and Ewen Cottage. Drake Cottage showed a very slight tendency to decline. These results are consistent with the earlier results with respect to Ewen Cottage which showed a tendency for these younger subjects to improve during the experimental period itself.

For Graves, however, it is an entirely new finding and a trend that did not show up until the follow-up test was given.

Analysis of Results on the Point Reward Score Sheet

The Point Reward Score Sheet (see Appendix B) was utilized in this experiment as the main tool by which subjects' progress was measured. It was felt that by analyzing some major areas of information regarding subjects' progress, that a better understanding of the phenomonology of the experiment could be achieved.

Table 17 is a summary of four major categories of information provided by the Point Reward Score Sheet.

Table 16

Direction and Magnitude of z Score Post to Follow-Up Means for Experimental Group Cottages Across Dependent Variables Where Significant Differences Occurred

Variable	Cottage Group	Mean
Poor Control	Graves	+.33
	Drake	-.17
	Ewen	+.90
Inability to Delay	Graves	+.41
	Drake	-.06
	Ewen	+.85

Table 17

Statistics in Relation to the Experimental Group's Achievement on the Point Reward Score Sheet

	Age of S. at Onset of Exp.	Ratio of Weeks Earning Weekly Rewards	% of Weeks Earning Weekly Rewards	Mean Weekly Score per S.
Graves				
S. 1	13.7	11/11	100.0	695.5
S. 2	17.7	10/11	90.9	667.7
S. 3	16.2	11/11	100.0	697.3
S. 4	15.2	2/5	40.0	580.0
S. 5	14.7	10/11	90.9	664.5
S. 6	14.0	8/10	80.0	648.0
S. 7	16.9	10/11	90.9	681.0
S. 8	16.1	7/11	63.6	638.2
Graves Total	16.5	69/81	85.2	659.1
Drake				
S. 9	16.4	7/12	58.3	636.7
S. 10	16.3	5/12	41.7	618.3
S. 11	15.5	11/12	91.7	668.8
S. 12	15.0	11/12	91.7	662.5
S. 13	15.0	9/12	75.0	657.9
S. 14	13.0	9/10	90.0	673.0
Drake Total	15.2	52/70	74.3	652.9
Ewen				
S. 15	14.0	3/12	25.0	611.3
S. 16	12.9	6/12	50.0	630.8
S. 17	13.2	0/12	0.0	608.3
S. 18	12.9	7/12	58.3	649.6
S. 19	15.2	1/9	11.1	593.9
S. 20	13.2	8/12	66.7	647.1
S. 21	13.1	0/9	0.0	587.2
Ewen Total	13.3	25/78	32.1	618.3
Grand Total		146/229	63.8	

Column One gives the age of each of the twenty-one subjects at the onset of the experiment. Column two indicates the number of weeks during the experiment that the subject earned the weekly reward of the $1.50 trip. This is presented in ratio form with the denominator representing the maximum number of weeks that the subject actually participated in the study, and the numerator representing the number of weekly rewards earned. There are several subjects that did not participate for all twelve weeks. This is due to either home visits, runaways, a lost point sheet, or, in the case of two subjects, starting the program one or more weeks late due to being a new arrival at Lawrence Hall. Sixteen of twenty-one subjects actually participated for the maximum twelve weeks, but for five of these in Graves Cottage, the point sheets were lost for one week, thus their total number of weeks is eleven rather than twelve.

The third column provides percentages for the ratios given in Column Two. The fourth column gives the mean weekly score per subject during the entire number of weeks that he participated in the experiment. Maximum score was 700 points per week.

There were marked age differences among the cottages in this study. The mean age for Graves was 15.6 (range 13.7-17.7) and for Drake was 15.2 (range 13.0-16.4). The mean age for Ewen, however, was 13.3 (range

12.9-15.2). Thus subjects in Ewen generally averaged nearly two years younger than the other fourteen subjects in the experiment.

This age difference becomes noteworthy when the overall results as presented above are analyzed. It was only in Ewen that positive results approaching a significant level were obtained on the variables evaluated during the experiment itself.

Age also would appear to be related to the subjects' ability to earn weekly rewards. Boys in Graves earned weekly rewards 85.2 percent of the time, while boys in Drake earned weekly rewards 74.3 percent of the time (column 3). Boys in Ewen, however, earned weekly rewards only 32.1 percent of the time. None of the boys in Ewen were able to earn weekly rewards more than 75 percent of the time whereas 11 of the 14 subjects in Drake and Graves earned weekly rewards at least 75 percent of the time.

The implication of this was the PRSS was easiest for subjects in the two older cottages and most difficult for Ewen boys. However, since Ewen boys did better than the other two groups on their Devereux scores, it would appear that the PRSS was sufficiently challenging for them, but perhaps too easy for subjects in the other two cottages.

Supporting this conclusion were the mean average

scores per week. For Graves, it was 659.1 and for Drake 652.9, but for Ewen it was 618.3 (column 4).

The implication of the information provided by this table seems to be that the experimental condition was really most appropriate and challenging for the younger subjects. The older boys mastered the system rather easily. Paradoxically, though they obtained more rewards, they probably put less effort into altering their behavior and this was reflected in the lack of improvement in their Devereux Scale scores.

Younger subjects were being reinforced more intermittently than consistently. However, this seems to have provided better incentive for them to continue to try and move upward. The methodological implications of these observations are discussed in Section V.

The final grand reward was earned by a total of eight subjects. The reward consisted of dinner and a trip to a professional basketball game. Earning it was based on achieving the weekly reward ten out of twelve possible weeks or the equivalent thereof. None of the subjects in Ewen cottage were able to earn this reward.

Results of the Staff Questionnaire

As an additional method of obtaining information, a questionnaire was distributed to all staff participating in the program. This "Special Program Questionnaire" was

designed to be answered by staff in a brief period of time. Appendix C presents the staff questionnaire which was used as well as the results in tabulated form.

For each of the ten questions, staff were asked to choose from among three or four statements as to the statement that best reflected their opinion of the program.

Of the twelve staff members in the three units who were involved in the program, responses were obtained from eleven. Night staff who were not really involved at all in the administration of the program could have responded but presumably chose not to, although this cannot be ascertained for certain since questionnaires were anonymous.

The first question inquired as to whether the weekly point totals obtained on the PRSS seemed to be an accurate reflection of the boys' actual functioning during that week. Ten of eleven respondents felt that the weekly totals was a "somewhat accurate reflection of their behavioral functioning on the unit."

Staff were questioned as to the effectiveness of the daily reinforcers. Three of eleven felt that the daily rewards were a very good method of motivating boys. Six of eleven felt that daily rewards had "some positive effects" and two felt that these daily rewards had no significant effects at all.

Staff were asked to compare daily and weekly rewards as to which was most effective. Nine selected daily rewards as most effective, one selected weekly, and one could not judge.

However, when weekly rewards were considered alone, ten of eleven staff felt that boys "sometimes improve their behavior to obtain these rewards." The eleventh felt that they were unaffected by these rewards.

Staff were asked to evaluate the degree to which they noticed change in the boys over the course of the study. Seven of eleven felt that there were "some changes" and the other four felt that there were "no changes."

Staff were asked to evaluate the usefulness of continuing this or other similar programs. Two staff felt it would be good to continue this identical program while six suggested developing "similar but even better" behavior modification systems. Three suggested sticking to the system that existed prior to the study.

Staff were asked to what degree the program was an extra burden to them in terms of adding to their overall work load. Ten of eleven felt the program entailed "some extra work . . . but not an unpleasant amount." The eleventh felt the program was a considerable burden.

Staff were asked to rate this experimenter on his ability to follow through with the program. Two of eleven

felt that he was "very reliable" in terms of following through. Five of eleven felt that he was "usually but not always reliable." Four of eleven felt that he was "occasionally reliable."

The final question asked staff to describe their overall impression of the entire program. Four of eleven listed it as "very beneficial," four of eleven felt it was "somewhat beneficial," and three of eleven felt it "had no effect" at all.

In summary, the impression which this questionnaire conveys is that staff working in the experimental condition felt that the program was effective to a moderate degree. They endorsed it, but they were reserved in their endorsement. Generally, where a few extremely favorable responses were recorded, these came from Ewen cottage, while the more moderate responses came from Graves and Drake. These findings are consistent with the findings in regard to the Devereux.

However, this questionnaire is a validational tool of a very different sort than the Devereux. It recorded the entirely subjective and very global impressions of staff. In the case of the Devereux, very subtle or moderate improvement would not necessarily show up as statistically significant, whereas this questionnaire did pick up staff's point of view that there was almost across

the board moderate improvement. Further implications of these findings are discussed in Section V.

Intelligence of Subjects

There was difficulty in obtaining I.Q.'s for each of the subjects in the population. Lawrence Hall does not routinely test new residents. They do rely on I.Q.'s obtained from the referral material, but this material is not always complete.

I.Q.'s were obtained for 14 of 21 subjects in the experimental group. The mean full scale I.Q. for this group was 87.06 with a standard deviation of 16.11. For the control group, I.Q.'s were obtained for 13 of 21 subjects. The mean full scale I.Q. was 89.92 and the standard deviation 20.36. A t test performed for unequal N's resulted in $t = .43$ which indicated that the two samples did not differ significantly. It was, therefore concluded that I.Q. was not a significant differentiating variable between experimental and control group subjects.

Also, it needs to be considered that these I.Q.'s were obtained from different sources and differ considerably in when they were obtained. Some were quite recent and others were two-four years old.

It would not be reasonable to compute I.Q.'s as a mean for cottage groups. This is because at least two and sometimes three I.Q.'s are missing from each cottage group of only six-eight subjects.

V. DISCUSSION

The major hypothesis of this study was that a behavior modification program could have a significant impact on measurable behavior of emotionally disturbed adolescent boys, in comparison to an untreated control group.

The hypothesis was supported to a certain extent, especially in regard to an apparent delayed improvement in target behaviors between the post test and the follow-up text sixty days later. From pre to post test only, there were no significant differences between control and experimental subjects. The results are thus felt to indicate a tendency to support the hypothesis, but a much weaker tendency than was hypothesized at the onset of the experiment.

Methodological Issues

It is felt that several problems in the methodology may have contributed to the discriminative power of this experiment being less conclusive than was generally anticipated. The first part of this discussion will attempt to analyze apparent methodological problems that arose in the study.

The implication of Table 17 (see p. 88) was

that the experimental procedure was really more appropriate for the younger (Ewen cottage) subjects. Many of the older boys were already functioning behaviorally at a level that made it unnecessary for them to substantially alter their behaviors in order to obtain maximum rewards from the system. This was in turn reflected in the lack of improvement in their Devereux Scale scores.

Seven of 21 boys, all in Graves and Drake cottages, started off getting very high scores daily and simply continued to do so throughout the 12 weeks. Since they were already functioning at the beginning of the program at a level that enabled them to achieve rewards immediately, there was no real need for them to improve these behaviors. The target behaviors were too easy for these boys.

The implication of the above data is that the Point Reward Score Sheet was not universally relevant to each subject. For some subjects, it was appropriate, for others either too difficult or too easy.

For example, consider item 10 on the PRSS, "Did he interrupt others when talking?" This behavior was derived for the PRSS from a specific item on the Devereux Scale related to hyperactivity. However, for most boys in the experiment, this behavior was never a significant

problem to begin with. For these boys this (and several other items) became an automatic score of "5" every day of the program. There was no need for them to modify any behavior in order to obtain the maximum number of points.

Other items were more relevant to a greater number of boys, for example, number 7, "Did he do his homework without resistance?" This item was a relevant and adequately challenging target behavior for most of the boys in the study.

In general, the PRSS was only effective with a small number of subjects, generally the younger ones. Target behaviors represented by its twenty items were simply not relevant to all boys. Boys for whom the behaviors were too easy were receiving constant rewards but were really not improving any behaviors. Boys for whom the project was too difficult (two boys never once received a weekly reward) never received initial or intermittent reinforcement which would have encouraged them to try harder. Only boys in the middle area who were receiving intermittant reinforcements based on their level of achievement, did well in the program.

In retrospect, it might have been better to create the Point Reward Score Sheet on a more individ-

ualized basis, based on observable target behaviors in the boys rather than basing it on Devereux criterion factors as was done. This would result in individualized point sheets for each boy. Daily points would be based on specific target behaviors that were relevant to a particular boy's problems. These programs could be revised and updated as often as was necessary with the target behaviors increasing in difficulty as the boy improved. This would insure that each subject was being judged on the basis of behavioral goals that were sufficiently challenging to him.

There are, however, disadvantages to this method. It is more subjective because staff are judging boys as an "n of 1" rather than all boys being evaluated on the same target behaviors, thus establishing norms. However, this problem can be partially ameliorated by having target behaviors described in strict behavioral terms, thereby minimizing the need for subjective judgments.

A second problem is that an individualized system requires much more work on the part of staff. When old behavioral contracts are mastered, new ones must be written. In addition, staff is making daily evaluations of many more behaviors because of the greater variety in the boys' programs. Each of eight boys in a given cottage may be being evaluated on 15-20 entirely different be-

haviors each day, behaviors which may or may not overlap with those that the other seven boys are being rated on. Thus the complexity of successfully administering the program increases considerably. If staff is motivated to put forth the required investment, the program might succeed, but, if not, it will probably fail. The whole issue of staff investment and morale in relation to this type of project is discussed later in this section. This issue, in turn, is related to the problem of designing applied research studies that meet criteria for sound experimental design, and also are practical to conduct. No matter how excellent a written experimental methodology may be, it is relatively useless unless real people are able and willing to implement it.

Problems of Reinforcers

In any successful behavior modification program, two basic criteria must be met. As discussed above, appropriate target behaviors must be devised. Secondly, reinforcements must be meaningful for subjects. It appears that reinforcers in the present study were highly appropriate for the subjects around thirteen years of age, but only somewhat appropriate for older subjects. The conclusion is that you cannot assume uniformity of reinforcements for your population. Even the fact that a somewhat diverse "menu of reinforcements" was provided during this experi-

ment appears to have been inadequate to appeal to many of the older boys. The implication for future research is that there needs to be more assessment of what is reinforcing for individuals within the target population. This assessment needs to be done prior to the experiment.

Obviously cost issues are related to issues of reinforcements. Costs of the study are discussed in Appendix E. One goal must be to find meaningful reinforcers that are not expensive beyond the bounds of the available budget. It is possible, however, that reinforcers for older boys are going to be more expensive than for younger subjects. Most sixteen year olds are simply not motivated on a daily basis by a 20 cent item. The younger the subject the greater the reinforcing power of inexpensive material items.

One alternative worth further consideration is the exploration of reinforcers that are not necessarily material. For example, forming an exclusive club or group for boys who earn maximum points. In this case, prestige might be reinforcing. Also, the opportunity to spend additional time with staff members can be very reinforcing to some boys. Inexpensive field trips can be considered as incentives if there is sufficient staff to conduct such programs.

It was observed throughout this study that daily and weekly rewards both seemed to be very rewarding to

the younger boys. They would constantly talk to the experimenter about the kinds of candy they preferred. When the experimenter brought a new selection of candy into the cottage, they would generally follow him into the office and pay a great deal of attention to these rewards.

Older boys, however, appeared fairly indifferent most of the time to the daily rewards. However, they did show interest throughout the experiment in weekly rewards. Weekly rewards were presented on Sunday afternoons and the subjects were obviously awaiting the arrival of the experimenter each Sunday. Although the option existed of taking the $1.50 reward in cash, boys almost universally chose to go to a drug store or restaurant with the experimenter. Younger boys in Ewen cottage were also very interested in and apparently motivated by the weekly rewards, although they did not obtain them as often.

It appears difficult to assess the effect of the final "grand reward" on the subjects. This reward involved dinner and a trip to a professional basketball game, a package worth about $9.00 per subject. It appeared during the course of the experiment that older subjects more frequently would discuss this grand prize. It did seem to be of some interest to them. Younger boys in general did not appear to be able to deal with the fact that this reward was far off in the future. The criteria on which it was to be obtained (earning 650 or more points 10

weeks out of 12), although carefully explained to all subjects, may have been more difficult for younger subjects to comprehend. Older subjects, on the other hand, were well aware of their progress toward this particular goal.

Another item worth discussing is the "Certificate of Good Behavior" (see Appendix D). This was presented each week after scores were tallied. Boys names were written in on the certificate and the staff distributed it to the winners. Subjects in Drake and Graves generally displayed little interest in this Certificate. Younger subjects, on the other hand, asked about it more frequently, and were seen to tape these up on the walls or doors of their rooms. Early in the experiment they would make inquiries to staff if they did not receive the certificate right away. In general, as the experiment continued the certificate seemed to diminish in interest to all subjects.

The main implication of all this for future researchers is that they need to strive to develop meaningful reinforcers for each participating subject. The uniform value of the reinforcers cannot be assumed, and reinforcers generally differ in value depending on the age of the target population. The daily menu of reinforcement utilized in this study was more meaningful for subjects whose age was fourteen or under. The weekly reinforce-

ments of $1.50 in food or goods seemed most reinforcing to the older subjects.

Staff Issues

An issue warranting elaboration is the relationship of staff morale to the success of a program such as this. The staff will, in most instances, be the persons directly responsible for the daily implementation of a behavior modification program. Traditionally, child care workers in residential facilities are underpaid and there is a high turnover rate. Morale is extremely variable. Child care workers are frequently not involved in administrative decision making and may be resentful of "professionals" who receive more money and work less hours. Presumably, where moral is low, there will be resistance to changes in the cottage system brought about by an outside researcher.

In this experiment, an attempt was made to minimize resistance by meeting with staff members, explaining the project to them, and attempting to enlist their cooperation. They were told that this overall system could provide them with some incentives to offer boys in their program for good behavior. Staff did seem impressed by the very tangible rewards that were being offered in the project. Staff at Lawrence Hall frequently do complain that there are not enough material benefits for

the boys, such as better allowances, more off campus trips, etc. This program would provide some additional rewards for boys, even if only temporarily.

It was probably also helpful in enlisting the co-operation of Lawrence Hall staff that several staff participating knew the experimenter from his previous position in the Agency. Thus, he was not as likely to be viewed as an outsider.

It was expected that if initial resistance to this program could be minimized in staff, that the continued participation of the staff would be related to their perception of whether the experimental procedure was relevant or beneficial to the work they were doing. This hypothesis was indeed supported in the overall impression conveyed to this experimenter.

For example, child care workers at Ewen Cottage seemed to decide rather quickly that this system was having a desirable effect on the boys. These staff members seemed to become more enthusiastic about the project. They went out of their way to give positive feedback to the experimenter whenever he brought reinforcements around. They seemed to perceive the experimental condition as a very useful addition to their program. In turn, they conveyed this attitude to the boys and probably spent more time than any other unit reminding boys of the reinforcements and desired behaviors. Later, at the end of the

project, they informed the experimenter that they were using some leftover Christmas money to continue the program in a modified form. This was probably the most significant testimony to their belief in the program's effectiveness.

In Graves Cottage, however, where staff apparently did not perceive much effect of the program on changing boys' behavior, staff interest in and support of the program seemed to wane somewhat as the weeks passed. The experimenter noticed several occasions when ratings were not done on a daily basis according to plan, but rather were being done every other day. One week all eight PRSS sheets were lost at the end of the week, so it was impossible to disburse weekly rewards. Graves subjects were scoring highest of any unit on the point sheets, but the staff, although they had initially been as enthusiastic as Ewen staff, did not seem to retain their initial motivation. The assumption is that this was due in part to their not perceiving meaningful behavior change leading to high scores, but simply that the items were too easy.

In Drake, a third kind of phenomenon took place. Drake staff displayed some initial resistance to the program, which was not displayed in the other two cottages. During the first two weeks they did not have the weekly

points added up when the experimenter came on Sunday to disburse rewards. They were also not following procedures when it came to disbursement of daily rewards. One staff member was saving up the daily rewards and giving them to the kids at the end of the week instead of at the end of each day as instructed.

Additional time was spent explaining the procedures to Drake staff. By the fourth week the program was running smoothly, and the staff there continued to be consistent for the duration of the study. The change was probably due to their perception of at least some of their boys taking a great interest in the attainment of weekly rewards. In spite of their own tendencies to be resistive, they were willing to acknowledge some apparently good effects of the program.

It is also felt that they were initially more resistant to the program because they are a more experienced staff who had worked together for a longer time. The experimenter was less known to them than he was in the other two cottages. They have a high sense of professional identity and probably tended to be skeptical of an outsider coming in and setting up a program that was time consuming for them to administer. In the long run, however, their competence as staff became evident and they were eventually able to operate the program very smoothly.

At the same time, they also probably perceived that some of the target behaviors were too easy for the boys and that the daily rewards were rather insufficient motivators. Thus their response was at the level of doing what was necessary, but without much additional enthusiasm.

It is apparent that staff attitude is a crucial factor in the successful implementation of this type of program. At Lawrence Hall, morale problems among staff did not seem significant at the time of the experiment. However, the perceived relevance of the program to the boys' problems did seem to have a significant effect on the staff's overall performance in carrying through on the program plan.

Treatment Milieu Issues

In any study of one segment of a total treatment program, the effects of the total program on the analysis of the particular segment must be considered.

In this study, the control group was not exposed to a systematic behavior modification system during the three month experimental period. However, during that period the control group was receiving all of the other usual components of the Lawrence Hall treatment program. Such components include family and individual therapy, remedial education, and vocational training, as well as the overall effect of the cottage milieu which is de-

signed to be therapeutic. Behavior modification is used in a less systematic fashion to the extent that rewards and privileges are given to boys who are displaying the best behavior. For example, allowance may be withheld from boys for acting out behavior. Boys doing well during the week may be given more passes in the evening and on the weekends.

Also, the experimental group was receiving the other benefits of Lawrence Hall's program even while this specialized program was being conducted.

Therefore, it is assumed that control subjects would improve to a certain degree simply as a result of the regular treatment process. This effect cannot really be controlled for in a design of this sort. Similarly a certain amount of any improvement measured in the experimental group is likely to be due to the effects of the regular treatment components.

However, this study was set up as a stringent test of a specific and systematic behavior modification program that was added to the regular Lawrence Hall program. The subjects' behavior was being shaped in the direction of very specific, measurable target behaviors that would not necessarily be greatly effected by the regular Lawrence Hall program. There-

fore, the above effect is not considered to be a significant design problem.

Validation Issues

It is difficult to assess, based on the present study, if the Devereux Scale is the best method of evaluating improvement in a short term behavior modification study.

Jesness (in Buros, 1972) evaluates the Devereux in the following manner:

> The Devereux Adolescent Behavior Rating Scale should fill a useful function in clinical situations as well as in research studies. Because of the nature of the items it does not appear that the scale will prove useful for making fine discriminations among normal children. However, as the scale's authors suggest, the instrument can be recommended for use with diagnosed groups of disturbed children, and as a help in identifying disturbed children (pp. 134-135).

The question which arises, however, is what is the most appropriate type of research for the Devereux? The present study utilized a 12 week experimental time framework. It also utilized 5 of the 15 Devereux factors. The population was clearly a clinical one. Devereux factors are generally rated on a scale of 5 down to 1, 5 being very frequent occurrences of a symptomatic behavior and 1 representing the behavior never occurring. Factors are made up of four to six individual items, and one total score is obtained for the total number of items making up

a factor. The question of discriminative power simply means that if there is change in behavior that is meaningful, will this type of numerical scale reflect this change?

Another important question is whether or not three months is too short of a time framework to realistically expect observable changes in problem behaviors on this scale. Basically, staff's subjective ratings of improvement during the experiment were a little more favorable than results generated by the Devereux. Does this imply that the Devereux failed to discriminate more moderate improvement. This does seem likely, because small numerically positive changes in ratings may very well not lead to statistically significant pre-post test differences. There would have to be fairly large numerical changes in at least two-thirds of the items making up a Devereux factor. Thus it was possible in this experiment to have a boy improve in some specific areas but not have this improvement show up on the Devereux Scale. Major improvements would certainly show, but minor ones could easily be masked by lack of improvement or regression on other items.

When the sixty day follow-up study was conducted, the Devereux did pick up some measurable significant differences for two of the three cottages. This would in-

dicate that the Devereux is likely to discriminate more adequately for longer term experiments.

In this particular study, evaluation of the discriminative power of the Devereux is made more difficult by the confounding variables of lack of appropriate target behaviors for some subjects and non-uniform reinforcers. Given the probable effect of these two methodological weaknesses, it is difficult to make a definitive conclusion regarding the Devereux.

One significant advantage of the Devereux is that staff biases of any sort are not as likely to show up in the ratings as they are in a more subjective staff questionnaire. This is because staff are naive as to which items on the Devereux are crucial and what the true meaning of the numerical ratings are in terms of the ultimate results of the experimental program.

The alternative to using the Devereux, if one were necessary, would be to find another means of validating behavior change. For disturbed adolescents in residential treatment, there does not appear to be much available of a standardized nature. Most often, other evaluations of adolescent behavior are based on the rater's perception of intrapsychic change, rather than on change in specific overt behaviors. Thus these instruments would probably not be useful in a short term behavior modification study.

If alternative standardized tests do not meet the

necessary criteria for evaluating specific behaviors, the main additional option available is to devise a rating scale specifically for a given experiment. Such a scale could presumably be constructed in behavioral units that are coordinated with the goals of the experiment. The rated items could be directly focused on the target behaviors. The problem here becomes one of trading specificity at the expense of established reliability and validity. The Devereux adequately meets standards of reliability and validity. It has been utilized with a large number of very specific clinical populations and norms for these groups have been established. A newly constructed scale may not match these standards. Thus, results can be questioned.

There is a great need however, for the construction of new tests relevant to improvement in residential settings. If a newly constructed scale appeared to generate some useful results, perhaps other researchers would go further in trying it out with other populations thus working for the attainment of validity.

There was a questionnaire utilized in this study which polled staff's overall attitudes toward the project. It was felt that this questionnaire provided some useful information. However, it does not qualify as a really useful tool in a thoroughly scientific sense. However, any study of behavioral change should perhaps contain a

similar type of questionnaire polling staff perceptions of change as an additional means of confirming the apparent results from the main instrument.

Effects of Experimental and Control Group Subject and Staff Interactions

The factor of interaction between the boys in the various cottages was not considered to be a variable significantly effecting the outcome of the experiment. Cottages units at Lawrence Hall operate pretty much autonomously. The only time that the subjects have significant sustained contact with each other is at school, where much of their time is structured.

Again, the level of sophistication of boys at Lawrence Hall is not such that they discuss treatment strategies with each other in any detail.

Observations of control group subjects did not indicate that they were at all aware of any significantly different program taking place in the experimental units. In fact, control as well as experimental subjects were typically self-centered and not overly concerned with what was happening to other boys outside their own cottage unit.

Staff function even more autonomously in their jobs than kids. They are assigned to and work at a specific cottage and do not have time to interact with staff from other cottages. Only in emergency situations do they have contact with outside staff.

Certainly staff in the experimental cottages were aware that a "special program" was taking place in their own unit and in some other units, but this did not appear to have any effect whatsoever on their effectiveness in administering the program. Other factors which did have some presumed effect have already been discussed.

Issues in Regard to the Follow-Up Study

The results of the analysis of variance which compared pre-test scores with scores obtained sixty days after the conclusion of the experiment showed that the experimental and control populations differed significantly on two of the main dependent variables. Most of this improvement occurred between the post-test and the follow-up and the main source of the improvement were Graves and Ewen Cottages. Drake Cottage did not show significant differences over the three tests.

It is difficult to be definitive in terms of the exact causes of this result. The trend for improvement was consistent for Ewen subjects, but for Graves subjects, it was a significantly different result. The implication is that Graves subjects were effected by the PRSS program, but that the effects of the program were not measurable at the time of the post-test. It could well be, however, that as a result of reinforcements being withdrawn at the termination of the experimental program, that there was

actually a corresponding increase on the part of some subjects in behavior that had initially resulted in reinforcement.

Related to this would be a phenomenon in which praise and staff appreciation of previous behavioral change became significantly reinforcing to subjects even after the cessation of the more tangible reinforcements. Staff enjoyed the changes that had been produced in the subjects and the subjects were continuing to receive gratification from this appreciation.

The possibility of a "historical" influence on the subjects during the sixty day follow-up interval cannot be overlooked.

This would be some sort of change in the structure of the program or in the staff pattern that would lead to a sudden improvement in boys' behavior. However, when this possibility was investigated by interviewing administrative and child care staff, it was found that no significant program changes had occurred. There was no staff turnover, and no other visible changes in the unit.

Thus one is left to surmise that a response to the withdrawal of the reinforcement of the program in some fashion actually stimulated further shaping towards the original target behaviors. It is unfortunate that this same effect did not occur in the Drake unit, because it

would certainly add some strength to this conclusion.

It seems obvious that the follow-up study raises further questions for investigation in terms of delayed effects of behavior modification programs, effects that may in fact show up after discharge from the treatment center.

SUMMARY

The present study was a description and analysis of a controlled behavior modification intervention experiment in a residential treatment center for emotionally disturbed boys.

Forty-two boys between the ages of 13 and 18 were given the Devereux Adolescent Behavior Rating Scale. Twenty-one boys were assigned to an untreated control group and twenty-one to the experimental group. Boys living in three cottages made up the experimental group while the control group consisted of subjects living in four cottages.

For the next three months, boys living in the experimental cottages received systematic daily and weekly reinforcements for appropriate behaviors. These behaviors were measured on a "Point Reward Score Sheet" where each subject could earn a maximum of 100 points per day. Subjects receiving at least 90 points per day would receive a daily reward consisting of an item valued at about 20 cents. Boys earning at least 650 points in a week were entitled to a weekly reward of a trip to a store or restaurant to purchase $1.50 in food or goods.

The Point Reward Score Sheet was designed to reinforce target behaviors drawn from the Devereux Scale. The

five main target behaviors were unethical behavior, defiant-resistive behavior, hyperactivity, poor emotional control, and inability to delay gratification. The raters for the point reward score sheet were the child care workers in the unit where the subjects resided. Child care workers received instructions in administering the Point Reward Score Sheet. One child care worker in each unit was also assigned to do the rating on the Devereux Scale and instruction was given in this as well.

Results indicated that the experimental group did not differ significantly on any of the five variables from the control group between the pre and post test. However, when a follow-up test was administered sixty days later, it was found that the experimental group subjects did differ significantly from control subjects on the variables of poor emotional control ($p < .01$) and inability to delay gratification ($p < .01$). Further analysis indicated that the source of this change was attributed to two of the three experimental group cottages. The third cottage showed no improvement.

In the youngest cottage, Ewen, the trend to improve was somewhat consistent during the experimental interval also, but for the other cottage the significant gains made during the follow-up phase were discussed in terms of various possible delayed effects of the experimental program.

Although the main hypothesis in the experiment was only partially supported, it was felt that the study provides some useful directions to move in in relation to evaluating behavior modification in residential treatment centers.

Two main areas that were discussed were the difficulties in providing reinforcers that were of uniform value to the population and differences in degree of difficulty of the items on the Point Reward Score Sheet for the subjects. Regarding the reinforcers, it was felt that younger subjects were more motivated in general by the reinforcements provided in the study. Older subjects enjoyed the weekly reinforcers but were not well motivated by the daily items.

Older subjects were also not challenged as much by the Point Reward Score Sheet. Many of the items were too easy and older subjects were able to obtain consistently high scores without having to modify any behaviors. The program was much more challenging for the younger cottage subjects.

The appropriateness of the Devereux Scale for an experiment such as this is also discussed. Some alternative validational tools are also suggested. This study did utilize a staff questionnaire at the very end of the program to get a subjective impression of how staff viewed

this project. Similar questionnaires could be considered as validational tools although there are drawbacks in terms of reliability and previously established validity.

The I.Q.'s of some subjects were analyzed although I.Q.'s for the total population were not available. It was not felt the I.Q. was a distinguishing variable between the experimental and control group populations.

The experiment is discussed in terms of providing a replicable design methodology for developing a body of knowledge regarding the usefulness of behavior modification as an effective therapeutic tool in residential treatment.

REFERENCES

Aichhorn, A. Wayward youth. New York: Viking Press, 1934.

Allerhand, M.; Weber, R.; and Haug, M. Adaptation and adaptability: The Bellefaire follow-up study. New York: Child Welfare League of America, 1966.

Bettelheim, B. Love is not enough. Glencoe: Free Press, 1950.

Birnbach, D. The skills of child care. In W. Schwartz and S. Zalba (Eds.), The practice of group work. New York: Columbia University Press, 1971.

Browne, C. Some problems of children's institutions in achieving maturity. Child Welfare, 1963, 42, 77-81.

Buros, O. K. Seventh mental measurements yearbook. Newark: Gryphon Press, 1972.

Butler, J., and Haigh, G. Changes in the relation between self-concepts and ideal concepts consequent upon client-centered counseling. In C. Rogers and R. Dymond, Psychotherapy and perconality change. Chicago: University of Chicago Press, 1954.

Campbell, D. T., and Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1970.

Chandler, M.; Greenspan, S.; and Barenboim, C. Assessment and training of role-taking and referential communication skills in institutionalized emotionally disturbed children. Developmental Psychology, 1974, 10, 546-553.

Coche, E., and Thomas, A. Evaluative research on a therapeutic community for adolescents. Journal of Youth and Adolescence, 1975, 4, 321-330.

Cochrane, R. The impact of a training school experience on the value systems of young offenders. British Journal of Criminology, 1974, 14, 336-344.

Cohen, J. Multiple regression analysis for the behavioral sciences. Psychological Bulletin, 1968, 70, 426-443.

Davids, A.; Ryan, R.; and Salvatore, P. Effectiveness of residential treatment for psychotic and disturbed children. American Orthopsychiatric Journal, 1968, 38, 469-475.

Dubois, J. Residential treatment follow-up: Review of literature and report of follow-up study of Ninety pre-adolescents. Mimeographed research report. Madame Vanier Children's Services: London, Ontario, 1972.

Durkin, R., and Durkin, A. Evaluating residential treatment programs for disturbed children. In M. Guttentage and E. Struening (Eds.), Handbook of evaluation research. (Vol. II.) Beverly Hills: Sage Publications, 1975.

Easson, W. The severely disturbed adolescent. New York: International Universities Press, 1969.

Eaton, L., and Menolascino, F. Psychotic reactions of childhood: A follow-up study. American Journal of Orthopsychiatry, 1967, 37, 521-529.

Eisenberg, L. The autistic child in adolescence. American Journal of Psychiatry, 1956, 112, 607-612.

Etzioni, A. Two approaches to organizational analysis: A critique and a suggestion. Administrative Science Quarterly, 1960, 5, 257-278.

Feinsilver, D., and Gunderson, J. Psychotherapy for schizophrenics--is it indicated? A review of the relevant literature. Schizophrenia Bulletin, 1972, 6, 11-23.

Finkelstein, N. Family participation in residential treatment. Child Welfare, 1974, 53, 570-575.

Garber, B. Follow-up study of hospitalized adolescents. New York: Bruner Mazel, 1972.

Glasser, W. Reality therapy. New York: Harper & Row, 1975.

Glickman, E. Child placement through clinically oriented casework. New York: Columbia University Press, 1957.

Goldenberg, I. Build me a mountain: Youth, poverty and the creation of new settings. Cambridge: MIT Press, 1971.

Gough, H. California psychological inventory manual. Palo Alto: Consulting Psychologists Press, 1957.

Herrera, E.; Lifson, B.; Hartmann, E.; and Solomon, M. A 10-year follow-up of 55 hospitalized adolescents. American Journal of Psychiatry, 1974, 131, 769-774.

Herstein, N. The challenge of evaluation in residential treatment. Child Welfare, 1975, 54, 141-152.

Herstein, N. Reflections on the primacy of the one-to-one model in residential treatment. Child Welfare, 1977, 56, 311-320.

Inglis, D. Authority and reality in residential treatment. Child Welfare, 1974, 43, 273-279.

Joint Commission on Mental Health of Children. Crisis in mental health: Challenge for the 1970's. New York: Harper & Row, 1970.

Kahn, M., and McFarland, J. A demographic and treatment evaluation study of institutionalized juvenile offenders. Journal of Community Psychology, 1973, 1, 282-284.

Kane, R., and Chambers, G. Seven-year follow-up of children hospitalized and discharged from a residential setting. American Journal of Psychiatry, 1961, 117, 1023-1027.

Krasner, L. Behavior therapy. Annual Review of Psychology, 1971, 22, 308-321.

Lander, J., and Schulman, R. The impact of the therapeutic milieu on the disturbed personality. Social Casework, 1960, 41, 403-408.

Lefkowitz, M. MMPI scores of juvenile delinquents adjusting to Institutionalization. Psychological Reports, 1966, 19, 911-914.

Levinson, R., and Kitchener, H. Treatment of delinquents: Comparison of four methods for assigning inmates to counselors. Journal of Consulting Psychology, 1966, 30, 364-365.

Levy, E. Long-term follow-up of former in-patients at the Children's Hospital of the Menninger Clinic. American Journal of Psychiatry, 1969, 125, 1633-1639.

Maluccio, A. Residential treatment of disturbed children: A study of service delivery. Child Welfare, 1974, 53, 225-235.

Maluccio, A., and Marlow, W. Residential treatment of emotionally disturbed children: A review of the literature. Social Service Review, 1976, 42, 230-250.

Matushima, J. Group work with emotionally disturbed children. Social Work, 1962, 7, 62-70.

Mayer, M. The role of residential treatment for children-introduction. American Journal of Orthopsychiatry, 1955, 25, 667-668.

Novotny, E., and Burstein, M. Public school adjustment of delinquent boys after release from a juvenile corrective facility. Journal of Youth and Adolescence, 1974, 3, 49-61.

Offer, D., and Howard, K. An empirical analysis of the Offer Self-Image Questionnaire for Adolescents. Archives of General Psychiatry, 1972, 27, 529-533.

Pappenfort, D.; Dinwoodie, A.; and Kilpatrick, D. Population of children's residential institutions in the United States. Chicago: Center for Urban Studies, 1968.

Persons, R. Psychological and behavioral change in delinquents following psychotherapy. Journal of Clinical Psychology, 1966, 22, 337-340.

Persons, R. Relationship between psychotherapy with institutionalized boys and subsequent community adjustment. Journal of Consulting Psychology, 1967, 31, 137-141.

Polsky, H., and Claster, D. The dynamics of residential

treatment: A social systems analysis. Chapel Hill: University of North Carolina Press, 1968.

Redl, F. The concept of a therapeutic milieu. American Journal of Orthospychiatry, 1959, 24, 721-736.

Roen, S., and Burns, A. Community adaptation schedule. New York: Behavioral Publications, 1968.

Rossman, P., and Knesper, D. The early phase of hospital treatment for disruptive adolescents. Journal of the American Academy of Child Psychiatry, 1976, 15, 693-708.

Rotter, J. Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs, 1966, 80, 103-111.

Sarason, I., and Ganzer, V. Anxiety, reinforcement, and experimental interactions in a free verbalization situation. Journal of Abnormal and Social Psychology, 1962, 65, 300-307.

Sarason, I., and Ganzer, V. Modeling and group discussion in the rehabilitation of juvenile delinquents. Journal of Consulting Psychology, 1973, 20, 442-449.

Scallon, R.; Vitale, J.; and Eschenauer, R. Behavior modification in a residence and school for adolescent boys: A team approach. Child Welfare, 1976, 5, 561-571.

Schulberg, H., and Baker, F. Program evaluation models and the implementation of research findings. American Journal of Public Health, 1968, 58, 17-33.

Spivack, G.; Haimes, P.; and Spotts, J. Devereus Adolescent Behavior Rating Scale Manual. Devon, Pa.: Devereux Foundation, 1967.

Taylor, D., and Alpert, S. Continuity and support following residential treatment. New York: Child Welfare League of America, 1973.

Tortorella, W. Personality and intellectual changes in delinquent girls following long-term institutional placement. Journal of Community Psychology, 1973, 1, 288-291.

Treffert, D. Child-adolescent unit in a psychiatric hospital: Five years later. Archives of General Psychiatry, 1969, 21, 745-752.

Trieschman, A.; Whittaker, J.; and Brendtro, L. The other 23 hours: Child care work with emotionally disturbed children in a therapeutic milieu. Chicago: Aldine-Atherton, 1969.

Truax, C.; Wargo, D.; and Silber, L. Effects of group psychotherapy with high accurate empathy and non-possessive warmth upon female institutionalized delinquents. Journal of Abnormal Psychology, 1966, 71, 267-274.

Warren, W. A study of adolescent psychiatric in-patients and the outcome six or more years later. Journal of Child Psychology and Psychiatry, 1965, 6, Part I, 1-17; Part II, 141-160.

APPENDIX A

DEVEREUX ADOLESCENT BEHAVIOR RATING SCALE

AND SAMPLE PROFILE WITH ANALYSIS

[129]

DEVEREUX ADOLESCENT BEHAVIOR (DAB) RATING SCALE*

George Spivack, Ph.D.
Jules Spotts, Ph.D
Peter E. Haimes, Ph.D.

Devereux Foundation Institute for Research and Training

Youngster's Name Fred L. Rater's Name Bob H.

Youngster's Sex Male Rater's Relationship to Child Staff Member

Youngster's Birthdate 5-11-63 Date of Rating 6-17-77

RATING GUIDE

1. Base rating on youngster's recent and current behavior.	Consider only the behavior of the youngster over the past two (2) weeks.
2. Compare the youngster with normal adolescents his age.	In most of the items, the standard for comparison should be the normal adolescent of the same age and sex.
3. Base rating on your own experience with the youngster.	Consider only your own impressions. As much as possible, ignore what others have said about the youngster, and their impressions.
4. Consider each question independently.	Make no effort to describe a consistent behavioral picture or personality. It is known that adolescents may display seemingly contradictory behavior.
5. Avoid interpretations of "unconscious" motives and feelings.	As much as possible, base ratings on outward behavior you actually observe. Do not try to interpret what might be going on in the youngster's mind.
6. Use extreme ratings whenever warranted.	Avoid tending to rate near the middle of all scales. Make use of the full range offered by the scales.
7. Rate each item quickly.	If you are unable to reach a decision, go on to the next item and come back later to those you skipped.
8. Rate every question.	Attempt to rate each item. If you have had no opportunity to observe the youngster in certain situations necessary for the rating (e.g., "sexual relations", etc.), circle the item number.

The preparation of this publication was supported in part by Research Grant No. 1379-P from The Vocational Rehabilitation Administration, U.S. Department of Health, Education and Welfare.

YOU ARE GOING TO RATE THE OVERT BEHAVIOR OF AN ADOLESCENT. FOR ITEMS 1-57, USE THE RATING SCALE BELOW. WRITE YOUR RATING (NUMBER) FOR EACH ITEM IN THE BOX TO THE LEFT OF THE ITEM NUMBER.

Very frequently	Often	Occasionally	Rarely	Never
5	4	3	2	1

COMPARED TO NORMAL ADOLESCENTS HIS AGE, HOW OFTEN DOES THE YOUNGSTER...

Rating	Item
3	1. Show an interest in violence, death, people in accidents (e.g., in what he reads, talks about, watches on TV...., etc.)?
2	2. Have social contact with peers of the opposite sex?
2	3. Have a fixed facial expression that lacks feeling?
2	4. Intentionally tell lies?
4	5. Wear clothes that are provocative (e.g., short skirts and/or tight sweaters for girls; tight trousers and and/or open shirts for boys)?
3	6. Seek out adults for attention?
5	7. Persist when told he cannot have something (e.g., nag, demand, repeatedly ask for it, etc.)?
1	8. Express the belief that he has committed some unpardonable act, that he is evil, or that he deserves severe punishment?
4	9. Mumble, shout out, or make unusual vocal noises?
2	10. Cheat (e.g., in games, or sports)?
3	11. Mechanically repeat certain words or phrases in a meaningless way?
1	12. Daydream?
3	13. Mechanically repeat what is said to him (i.e., echolalia)?
2	14. Put inedible, unhealthy, or even dangerous things in his mouth (e.g., paper, wood, dirt, pins, garbage, etc.)?
2	15. Blame or condemn himself for things that happen to him?
4	16. Look puzzled or confused by things happening around him?
3	17. Get easily upset by peers (e.g., when pushed, teased, etc.)? (By peers is meant youngsters his own age, excluding brothers and sisters.)
5	18. Resist or refuse doing what is asked of him, or display a negative attitude?
4	19. Display odd facial grimaces, strange postures, or odd movements (e.g., hitting or biting himself, senseless or magical movements of the fingers, arms, legs or head, etc.)?
3	20. Tend to cling to adults (e.g., want to sit next to them, be around them a lot, etc.)?
4	21. Act bossy or domineering with other youngsters?
4	22. Express anger in a poorly controlled fashion?
4	23. Tend to be loud and boisterous?
4	24. Rock back and forth while sitting or standing?

Very frequently	Often	Occasionally	Rarely	Never
5	4	3	2	1

Rating	Item
4	25. Speak in a way that is disconnected, incoherent or not sensible (i.e., disregard speech handicaps and focus on the quality of the thought expressed)?
4	26. Express anger?
3	27. Exhibit interest in sex, through action or what he says?
3	28. Brag or act boastfully?
4	29. Walk around oblivious to what is going on around him (e.g., seem wrapped up in his own thoughts)?
2	30. Express the belief that others influence or control his thoughts (even though this isn't true)?
5	31. Appear overactive and constantly moving about?
1	32. Express gradiose ideas about himself which are extremely strange (e.g., that he has unusual or fantastic power over others, or things, that he is an extremely important person, etc.)?
3	33. Seem elated or high in mood?
1	34. Use his name rather than the word "I" when referring to himself in conversation (e.g., John went upstairs to get his coat)?
1	35. Substitute, confuse or misuse pronouns in conversation (e.g., use the pronoun "he" when referring to himself, confuse the pronouns "you" and "I", etc.)?
4	36. Tease or bully other youngsters? (Excluding brothers and sisters.)
1	37. Report hearing voices or other hallucinations?
4	38. Resent being told what to do?
4	39. Seek out adult approval and praise for what he has done?
4	40. Do what he wants to even when told he shouldn't (act defiant)?
2	41. Take things that do not belong to him (steal)?
1	42. Tell you things from his imagination as though they were really true?
5	43. Talk rapidly or hurriedly?
4	44. React with immediate anger or upset if he has difficulty mastering or learning something?
2	45. Make up his own words or use common words in such a peculiar way that it is difficult to understand what he means?

Very frequently	Often	Occasionally	Rarely	Never
5	4	3	2	1

Rating	Item
5	46. Act before he thinks (i.e., is impulsive)?
4	47. Do everything with boundless energy?
4	48. Get very upset or overemotional if things don't go his way?
2	49. Express depressed or despairing thoughts (e.g., express lack of hope, feelings of discouragement, that he expects the worst, no sense trying, etc.)?
4	50. Seek out adult help in doing things?
4	51. Insist on doing things his way?
3	52. Shut out sounds by lifting his shoulders to cover his ears, or putting his fingers in his ears?
3	53. Have a blank stare or far away look in his eyes?
2	54. Express the belief that people are against him (e.g., say that others pick on him, do not like him, talk about him behind his back, etc.)?
1	55. Express the belief that certain people are plotting or conspiring against him (e.g., secret police, criminals, international spies, etc.)?
1	56. Say that his body is diseased, distorted, or that his internal organs are rotted or missing?
1	57. Say that certain external forces (e.g., machines, electronic devices) are influencing or controlling his behavior and thinking?

FOR ITEMS 58-84, USE THE RATING SCALE BELOW:

Extremely	Markedly	Distinctly	Quite a bit	Moderately	A little	Very slightly	Not at all
8	7	6	5	4	3	2	1

TO WHAT DEGREE IS THE YOUNGSTER...

Rating	Item
2	58. Afraid of getting hurt in physical activities (e.g., climbing, roughhousing, sports, etc.)?
2	59. Preoccupied with compulsive acts he recognizes as unreasonable, but cannot stop himself from doing (e.g., touching, counting, certain acts or routines, etc.)?
2	60. Obsessed or preoccupied with ideas he worries or talks a lot about?
6	61. Impatient and unable to wait for things?
2	62. Unemotional - rarely shows feelings?

Extremely		Markedly		Distinctly	Quite a bit	Moderately	A little	Very slightly	Not at all
8		7		6	5	4	3	2	1

Rating	Item
3	63. Prone to avoid competition with peers?
3	64. Withdrawn from others his age (i.e., avoids social contacts, remains alone or isolated)?
4	65. Anxious or overconcerned about the future?
3	66. Boycrazy (for girls) or girlcrazy (for boys)?
2	67. Unaware of how adults feel toward him?
1	68. Lacking in muscle tone (e.g., when you feel his muscles they seem soft and doughy)?
5	69. Changeable or variable in mood or emotional state?
2	70. Physically weak?
2	71. Sneaky or underhanded in much of what he does?
6	72. Bossed or dominated by peers?
4	73. Poorly coordinated physically (e.g., clumsy or awkward in gross body movements, or in doing things with hands or fingers, etc.)?
2	74. Prone to tire quickly or have low endurance?
7	75. Prone to keep his distance or reserve with adults?
5	76. Unpredictable in his behavior?
2	77. Preoccupied with cosmetics (e.g., eye shadow, rouge; after shave lotion, hair tonic, etc)?
6	78. Unable to concentrate (e.g., jumps from one thing to another while talking or doing things, easily distracted in what he is doing by what others are doing around him, etc.)?
5	79. A fringe participant in peer social activities?
2	80. Timid or shy (i.e., will not "venture" out to try something new)?
7	81. Prone to hit or physically threaten peers?
6	82. Talkative?
7	83. Easily overexcited?
1	84. One whose contacts with peers of the opposite sex must be supervised?

DEVEREUX ADOLESCENT BEHAVIOR (DAB) RATING SCALE*

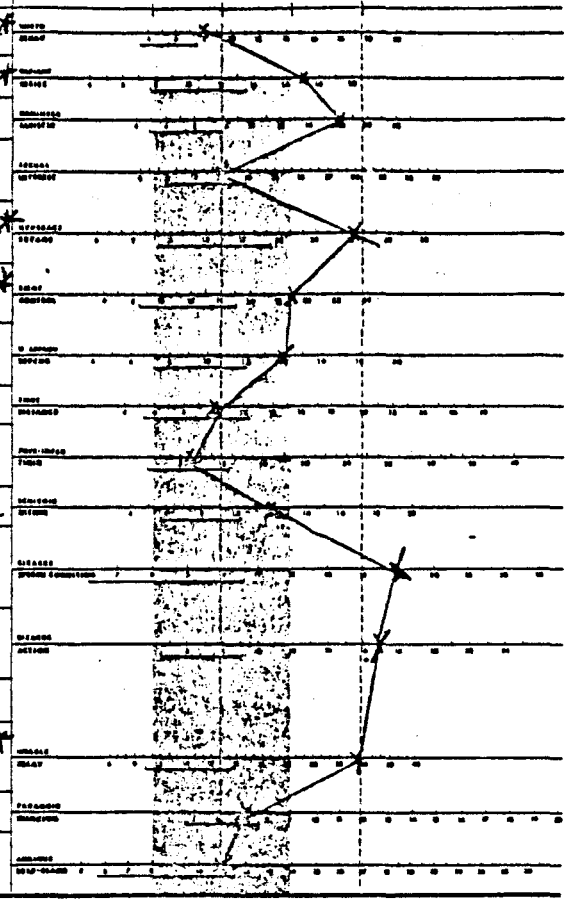
(DAB Profile)

George Spivack, Ph.D., Jules Spotts, Ph.D., Peter E. Haimes, Ph.D.

Behavior Factors	Factor Item Raw Scores	Sum Raw Sc.
1. Unethical	lies 4 2 41 2 steals cheat 10 2 71 2 sneaky	8 *
2. Defiant Resistive	neg. 18 5 40 4 defy resent 38 4 51 4 insist	17 *
3. Domineer Sadistic	boss 21 4 36 4 tease brag 28 3 81 4 phys. agg.	18
4. Hetero Sexual Interest	social 3 2 66 3 "crazy" clothes 5 4 77 3 cosmetic. interest 27 3 84 1 need support	15
5. Hyper-Activity Expansive	loud 23 4 43 5 talk fast hyper. 31 5 47 4 energ'c elated 33 3 82 6 illusive.	27 *
6. Poor Emotional Control	poor upset 17 2 44 4 im'd. mood no contr. 22 4 48 4 upset silly mad 26 4	19 *
7. Needs Approval Dependency	needs attn. 6 3 39 4 needs approval cling 20 3 50 4 help	14
8. Emotional Distance	face 3 2 67 2 unaware unemot. 62 2 75 4 adults	10
9. Physical Inferiority Timidity	fears hurt 58 2 74 3 tense tense 68 1 80 2 timid weak 70 2	9
10. Schizoid Withdrawal	day dreams 12 1 29 4 oblivious confuse. 16 4 53 3 stare	12
11. Bizarre Speech and Cognition	stereo. 11 3 35 1 pronoun. echol. 13 3 37 1 halluc. incoh. 25 4 45 2 neologs. l-norms 34 1	15
12. Bizarre Action	vocal 9 4 24 5 rock inadeb. 14 2 52 3 no. hour stereo. 19 4	17
Rational Clusters	**Cluster Item Raw Scores**	
1. Inability to Delay	neg 7 5 69 5 mood impulse. 46 5 76 5 unprod. impat. 61 8 83 7 excit.	35 *
2. Paranoid Thinking	viol. 1 3 32 1 grand. influ. 30 3 42 1 imag.	7
3. Anxious Self-Blame	sin 8 1 60 2 obsess self blame 15 3 65 4 ens. fut'r. despair 49 2	11

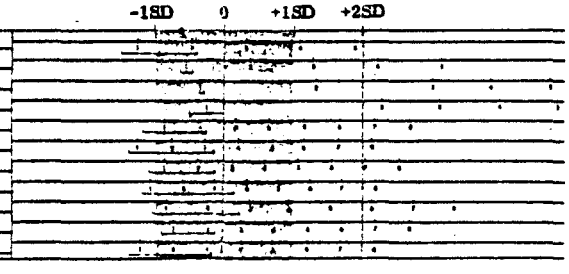
RAW SCORES IN STANDARD SCORE UNITS

-1SD 0 +1SD +2SD



Additional Items

54 persec
55 plots
56 body
57 influ
59 comp. act
63 avd. comp
64 withdrawn
79 fringe
72 bossed
73 coord
78 distract



(Youngster's Last Name) _____

(First Name) _____

Birth Date: _____

Date of Rating: _____

Age: _____ Sex: _____ IQ _____

Rater's Name _____ Relationship to Youngster _____

The sample profile provided here is a rating done by a child care worker on a fourteen year old resident of Lawrence Hall.

Fred was placed in Lawrence Hall after a year at Read Zone Center adolescent treatment unit. He has been diagnosed as schizophrenic and hyperactive and currently takes fairly large doses of Mellaril. His family background is very chaotic. He is the offspring of his mother's being raped while hitchhiking across country. The mother herself is an extremely impulsive and unstable woman who has been married several times. She appears highly schizophrenegenic. Fred was placed at Read Zone center two years ago after sexually molesting his younger sister. The sister is now also in placement at a State facility.

At Lawrence Hall, at the time of testing on the Devereux Scale, Fred was extremely uncontrollable. He was virtually unreachable in terms of either individual therapy or milieu approaches. His attention span was so short and his behavior so bizarre, that serious questions were raised as to his appropriateness for an open setting. (Subsequently, he has settled down somewhat, although progress is still extremely slow.)

The Devereux Profile confirms the diagnostic impressions. Fred's scores were indicative of considerable pathology on Factors 2, 3, 5, 11, and 12 of the profile, and also on Rational Cluster 1. He is seen by the rater

as resistant, sadistic, hyperactive, bizarre in terms of speech and action, and as being unable to delay gratification. In addition, he is abnormally poor in the area of emotional control and need for approval. In general, this is a highly pathological profile it reflects accurately many of the features of Fred's behavior that had been observed by staff at the time of testing.

In terms of the specific factors to be focused on in this particular experiment (marked with an asterisk), Fred was an extreme range on four of the five. This means that there would be considerable room for improvement on his part if an effective behavior modification program were to be introduced.

APPENDIX B

SAMPLE POINT REWARD SCORE SHEET FOR AN EXPERIMENTAL GROUP RESIDENT DURING THE SECOND WEEK OF THE PROGRAM

[138]

POINT REWARD SCORE SHEET

Instructions: Give either zero or five points for each of the 20 items listed. Five points are given for the most successful performance on the given item.

WEEK OF Oct 7TH-13TH

Name ROBERT F.

Rater JOSEPH W.

	Sun.	Mon.	Tues	Weds	Thu.	Fri.	Sat.
1. Did he tell any lies today? (5)	5	5	0	5	5	5	5
2. If there were games or sports, did he cheat? (5)	5	5	5	5	5	5	5
3. Was he caught stealing? (5)	5	5	5	5	5	5	5
4. Was he evasive or sneaky in any way? (5)	5	5	0	5	0	5	5
5. Did he do his regular cottage chores? (5)	5	5	5	5	5	5	5
6. Did he go to bed on time without resistance? (5)	5	5	5	5	5	5	5
7. Did he do his homework without resistance. (5)	5	5	5	5	5	5	5
8. Did he obey the overall rules of the program? (5)	5	5	0	0	5	5	5
9. Did he talk too fast to be understood. (5)	5	5	5	0	0	5	5
10. Did he interrupt others when talking? (5)	0	5	5	0	0	5	5
11. Was he unable to stay in one place to do an activity? (5)	5	5	5	5	0	5	5
12. Was he hyperactive to the point of disrupting the program? (5)	5	5	5	5	5	5	5

	Sun.	Mon.	Tue.	Wed.	Thu.	Fri.	Sat.
13. Did he have a verbal argument with any other kids? (5)	5	5	5	5	5	5	5
14. Did he have a temper tantrum if he failed at a task? (5)	5	5	5	5	5	5	5
15. Did he engage in a physical fight or need physical restraint? (5)	5	5	5	5	5	5	5
16. Swear at any staff members? (5)	5	5	5	5	5	5	5
17. Was he unable to wait when asked to wait? (5)	5	5	5	5	5	5	5
18. Did he persist when told he could not do something? (5)	5	5	5	5	5	5	5
19. Did he go through extreme mood changes requiring staff attention? (5)	5	5	5	5	5	5	5
20. Did he carry out a harmful or aggressive behavior without thinking of the consequences? (5)	5	5	5	5	5	5	5
Daily Total:	95	100	85	85	80	100	100

Weekly Total: 645

APPENDIX C

THE SPECIAL PROGRAM QUESTIONNAIRE WITH RESULTS OF THE SURVEY TABULATED

Special Program Questionnaire

Unit_____

This questionnaire concerns the special point reward system program that recently took place on your unit. The purpose of the questionnaire is to obtain your impression of the overall usefulness of this program. Please spend a few minutes on it as your opinion is important to evaluating this program.

Please put an X in the blank next to the item that best represents your opinion.

1. The point totals obtained by the boys on the special program in a given week usually was

   _____ a) a very accurate reflection of their behavior on the unit.

   10 b) a somewhat accurate reflection of their behavioral functioning on the unit.

   1 c) did not seem to correlate well with their behavior on the unit

   _____ d) was entirely misrepresenttative of their behavior on the unit.

2. As you are aware, boys received candy or gum on a daily basis as a reward for good behavior. Giving them daily rewards

   3/6 a) seems to be a very good method of motivating them.

   _____ b) seemed to have some positive effects on them

   2 c) seemed to have no effect on them.

   _____ d) seemed to have a negative effect on them.

3. In the program, two methods of reinforcement were used, weekly and daily. Of the two which seemed most effective?

   1 a) weekly.

   9 b) daily.

   1 c) could not judge.

4. Over the course of the 12 weeks of this program I have noticed

_____ a) very great positive changes in the behavior of the boys.

7 b) some changes in the behavior of the boys.

4 c) no changes in the behavior of the boys.

_____ d) primarily negative changes in the behavior of the boys.

5. With regard to the daily reinforcements, the kids in the unit generally seemed to

1/10 a) look forward to the rewards and shape their behavior to obtain them.

_____ b) sometimes improve their behavior to obtain rewards.

_____ c) never showed much interest in obtaining these rewards.

6. With regard to the weekly reinforcements, the kids in the unit generally seemed to

_____ a) look forward to these rewards and shape their behavior to obtain them.

10 b) sometimes improve their behavior to obtain these rewards.

1 c) never showed much interest in changing their behavior to obtain these rewards.

7. My overall impression would be that Lawrence Hall School for Boys should

2 a) continue this identical program if at all possible.

6 b) continue to develop similar but even better behavior modification programs.

3 c) stick to the basic system that now exists (prior to the special program.)

_____ d) discontinue these and other similar programs.

8. With regard to my overall work load as a child care worker, this special 12 week program

_____ a) did not really increase my owrkload at all.

10 b) was some extra work for me but not an unpleasant amount.

1 c) was a considerable burden on my being able to carry out other responsibilities on the job.

9. The experimenter in this program was

2 a) very reliable about bringing in the reinforcements and following through on the program

5 b) usually but not always reliable about bringing in reinforcements and following through on the program.

4

_____ c) occasionally reliable about these things.

_____ d) totally unreliable.

10. My overall impression of this special projects was that it was

4 a) very beneficial to this unit.

4 b) somewhat beneficial to this unit.

3 c) had no effect on the unit program.

_____ d) had a negative effect on the unit.

Below, please feel free to ad any further comments about this program and your reactions to it.

APPENDIX D

THE GOOD BEHAVIOR AWARD

GOOD BEHAVIOR

AWARD

is presented to

_____

IN RECOGNITION OF HAVING EARNED
650 OR MORE POINTS

WEEK OF: _____

SIGNED: _____

APPENDIX E

COSTS OF THE STUDY

The experimenter paid the total cost of the experiment. This cost was very close to $550.00. Of this, $260.00 was spent on daily reinforcers, $220.00 on weekly reinforcers, and an additional $70.00 on the final grand reinforcement for those subjects that earned it.

This breaks down to a cost per boy for the entire experiment of $26.19. The weekly cost per boy was $2.18. If the project were to have continued for a year, the cost per boy would be roughly $105.00.

Lawrence Hall's annual budget does not include a category for behavior modification reinforcers. The annual cost per boy at present is $55.00 per day or $20,075.00 annually. Staff salaries acount for the biggest portion of this, with food costs and plant upkeep also contributing significantly.

Considered from an overall cost of over $20,000.00 per year, $105.00 for behavior modification reinforcers would not seem to be a large added expense, especially if such reinforcers do have a significant therapeutic effect. Of course, this experiment has not entirely clarified the issue of how beneficial this type of program is. Many issues will remain for further investigation.

If a program such as this were to be adopted in a residential facility, there would be additional costs to consider, especially the time allotted for a professional

to administer the program. In this experiment, the researcher's time was free, but effective organization of behavior modification programs on a yearly basis might require a half or even full time psychologist. Of course, once programs are designed and implemented, paraprofessionals can continue to operate them.

It seems likely based on the results of this study that effective reinforcers for older subjects would turn out to be more expensive on the average than for younger subjects.

APPENDIX F

LETTER TO CHILD CARE STAFF

January 24, 1979

Dear _____

In the last several months you and several other staff members have done two ratings of Lawrence Hall boys using the Devereux Adolescent Behavior Rating Scale.

I am very appreciative of the time you have put into this and I hope it will contribute to the overall improvement of the program at Lawrence Hall. As it is a Dissertation research project I will be making the results available to interested staff as soon as possible.

Some of you will be asked to make one more rating on March 1st. I hope this will not inconvenience you too much.

Again, thank you very much for your help and cooperation.

Sincerely,

Richard Stern

cc: Gene B. Meier,
Executive Director

APPROVAL SHEET

The dissertation submitted by Richard A. Stern has been read and approved by the following committee:

Dr. Frank Kobler, Director
Professor, Psychology, Loyola

Dr. Roderick Pugh
Professor, Psychology, Loyola

Dr. Alan De Wolfe
Professor, Psychology, Loyola

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval by the Committee with reference to content and form.

The dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psychology.

June 19, 1979
Date

Frank Kobler
Director's Signature