This is to certify that the

thesis entitled

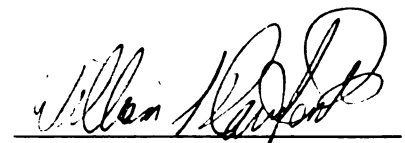
DEVELOPMENT OF A BEHAVIORAL MEASURE FOR SELECTION OF COLLEGE STUDENT VOLUNTEERS

presented by

CHRISTINA M. MITCHELL

has been accepted towards fulfillment
of the requirements for

Masters degree in Psychology



Major professor

Date 2-16-79

O-7639

DEVELOPMENT OF A BEHAVIORAL MEASURE FOR SELECTION OF COLLEGE STUDENT VOLUNTEERS

By

Christina M. Mitchell

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Department of Psychology

1979

ABSTRACT

DEVELOPMENT OF A BEHAVIORAL MEASURE FOR SELECTION OF COLLEGE STUDENT VOLUNTEERS

By

Christina M. Mitchell

The selection process is one which occurs in any program using nonprofessionals as change agents. Yet, there have been no clearly successful selection techniques created. Usually, traditional methods such as self-selection, interviews or paper and pencil measures are used which have little predictive validity. In this study, a behavioral measure was created using representatives of the target population role-playing situations typical to the intervention process. Creation of this measure is delineated; tests of its reliability and generalizability are reported; interrelationships with such measures as a personality measure, process data, in-class ratings by self, peers and supervisors are examined. Outcome, defined as success or failure of the client, is then used as a prediction criterion in the building of a selection battery through the use of discriminant function analyses. Implications for selection techniques for naturalistic interventions are discussed.

ACKNOWLEDGMENTS

This study was made possible by an outstanding amount of cooperation among many talented people. The support, guidance and prodding of Bill Davidson was invaluable to me; committee members Neal Schmitt and Charles Johnson invariably offered their time, expertise and concern, which I sincerely appreciate as well. All of the graduate students working on the Adolescent Diversion Project were extremely generous and flexible with their time; I would not have had access to as much data from such varied areas without all of them. A special thanks goes to Ricki Kantrowitz, who struggled with me through Student Data and Project hassles. And, of course, my love goes to my friends who gave me confidence, encouragement and respect, from both near and far.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The last decade has brought forth increasing debate concerning the area of juvenile delinquency. When compared with crime rates in 1960, statistics show an increase in the frequency of all crimes of violence of 190% (Federal Bureau of Investigation, 1973). If looking only at juvenile crime, approximately 34% of crimes going through law enforcement agencies involved persons 17 and under (Federal Bureau of Investigation, 1973). Thus, extrapolating from these figures, there has been an increase in reported juvenile crime.

Much of the debate concerns the current strategies in dealing with the juvenile offender. The first criticism, stemming from the Supreme Court decision concerning the case of Gault (1967), alleged that in the name of desire that the youth receive the best treatment possible, the courts were suspending or ignoring the constitutional rights of the youth. A second criticism focused upon the malignant correctional and treatment facilities for youth, which appeared to be doing more harm than benefit (Irwin, 1974). The third general criticism centered around the courts' determination to view the problems of delinquency as those of the apprehended and convicted youth alone, as opposed to those of youth and society in general (Erickson, 1973).

The turmoil present in the area of juvenile delinquency and its "treatment" led to several policy changes. One encompassed moving the focus of programs away from the malignant institutions and toward community-based intervention (Empey, 1967). With this recommendation came the concept of "diverting" youth from the justice system altogether. This suggestion stemmed from the idea that "the juvenile courts had been ineffective in preventing or correcting delinquency, and fiscal projections indicated that there were not likely to be the resources available to substantially upgrade the court's functioning" (Seidman, Rappaport, Davidson & Linney, in press). From this line of thinking arose the Youth Service Bureau concept, in which professionals attempted to deal with juvenile offenders via diversion. The literature on professional diversion attempts is at best weak, with little or no data concerning the impact of the programs on the individual and the systems involved (Lewis & Davidson, Note 1); however, it would appear that professional diversion has not provided an effective method of dealing with adolescent offenders.

The apparent failure of professional diversion has led to a second recommendation: the utilization of nonprofessional and volunteer staffing of diversion projects. There are several reasons for this suggestion. First, professional staffing shortages have been caused by the intra-individual approach taken by court workers. A similar problem has plagued the mental health field in general (Albee, 1968; Gruver, 1971). Also, present mental health ideology has prevented some subpopulations needing help--

e.g., alcoholics or juveniles--from receiving help, due to the fact that professional contact with them has proved to be fruitless. Of special interest to this study is Levitt's (1971) research showing that "conventional psychotherapy methods appear to be least effective with delinquents. The reported improvement rate is more than a standard deviation below the mean for all treated cases" (p. 484). Finally, the effectiveness of nonprofessionals over their professional counterparts has been shown with some populations. Gruver (1971) and Karlsruher (1974) have both substantiated that nonprofessionals were often more effective than professionals, thus suggesting that the use of nonprofessionals be expanded.

Gruver further expounded on this issue: "The fact that there is a reduction of social distance between college students and those seeking help may facilitate the establishment of a working relationship" (p. 113). On these grounds, Gruver purported that college students could perhaps be most effective working with other college students, adolescents and children.

With these issues in mind, the Adolescent Diversion Project at Michigan State University was funded by National Institute of Mental Health to examine nonprofessional diversion of juvenile offenders, based upon previous research (Davidson, Seidman, Rappaport, Berck, Rapp, Rhodes, & Herring, 1977). The issues involved with nonprofessional programs such as these are numerous: what kind of training is most effective, what type of intervention produces the most positive outcome, what is involved in the process of intervention, and, basic to all of these, <u>who is the</u>

most effective intervener. The decision of whom to select for nonprofessional programs in general is an issue of utmost importance, yet is one which is often treated speciously in the literature. Not only is the strategy rarely reported or researched, but the complete decision process is often unclarified and seemingly unimportant to program directors.

In this proposal, general selection issues will be briefly mentioned, followed by discussion of general decision theory in order to determine what would be a "good" selection measure(s) and a sound decision strategy for nonprofessional programs. Next, due to a dearth of literature on college nonprofessional diversion programs, literature concerning programs utilizing collece nonprofessionals with various target populations will be examined in light of their reported selection methods. Finally, a case will be made for the creation of a new, behavioral measure to be examined as a selection measure for nonprofessionals.

Selection and Prediction

Types of Selection Decisions

Facing any program director who has a selection decision facing her/him are four types of decision categories. If all applicants are to be selected and if there is only one type of "treatment" (used broadly here to refer to jobs, training programs, or any type of "categorizing" situation) available, the decision is termed one of acceptance. If some of the applicants can be rejected and if there is only one type of treatment to be assigned,

the decision is one of selection. If some of the applicants can be rejected and there are two or more different treatments available, one is faced with a multiple selection. Finally, if all of the applicants are to be accepted and there are two or more different treatments open, the decision is one of classification (Wiggins, 1973).

In addition to the options of rejection and multiple treatments, a decision maker must be aware of where s/he is going to place the ultimate value of a selection decision--will it be looked at from the viewpoint of the institution which is selecting, or from the viewpoint of the individual facing the selection? Also, the administrator can either choose a single-stage selection procedure, where the initial decision is the final one, or a multi-stage selection procedure, where decisions are made at various points concerning whether to reject an applicant or to gather more information on her/him (Wiggins).

Thus, before a program director can create a selection battery for the selection of nonprofessionals, it is important that s/he have a clear picture of the selection process in mind, for a selection battery well-suited for a single-stage selection procedure might be worthless if the decision is actually to be made in a multi-stage manner. As will be seen in later sections, although most authors rarely delineate their procedures precisely, it appears that most nonprofessional programs are concerned with institutional single-stage selection decisions. Thus, this

proposal will focus primarily on this type of decision and the creation of selection batteries for it.

Creation of a Selection Battery

In the prediction paradigm, psycholgoical "test" responses are "measurements which serve as input to a predictive system which combines data in such a way as to produce estimates of future criterion behavior" (Wiggins, p. 122). In this instance, the term "test" takes on a broader connotation than the common idea conjured by the word; it is more akin to Cronbach's (1970) definition--"A systematic procedure for observing a person's behavior and describing it with the aid of a numerical scale or a category-system" (p. 26). Thus, one must be knowledgeable about the data-collection component of the paradigm--the measurements--before examining the manner in which the measurements will be combined in order to produce a prediction of criterion behavior.

When a researcher is creating a battery of measures for selection, s/he is looking for a grouping which will order the applicants on their probabilities of success in that treatment. Thus, s/he is looking for tests with high correlations with the criterion performance scores (or a combination of criterion performance scores) and yet which have low correlations among themselves. These scores are compiled by a regression equation in order to produce a single prediction score, represented by

$$y' = b_1x_1 + b_2x_2 + \ldots + b_nx_n$$

where

y' - predicted score
b_i = estimated beta weight
x_i = score on the ith measure
n = total number of measures in the battery.

Thus, according to Wiggins, the best selection battery is one which "yields the highest correlation between predicted and obtained criterion scores, $r_{yy'}$, and hence yields the fewest errors in selection" (p. 233).

This particular model allows for compensatory scores--if a person scores low on one measure, s/he may be able to compensate for that low score by having a high score on another measure. There are some situations in which this model is not applicable--an applicant may need at least a minimum level of a certain trait before being considered at all. Wiggins cites the example of a pilot's scoring extremely high on a measure of pilot aptitude, but scoring extremely low on a visual acuity test. In this case, it is important that a pilot trainee have a minimum visual acuity before consideration of her/his pilot aptitude.

Again, although there are two options when creating a selection battery--multiple regression or multiple cut-off levels--nonprofessional literature has concerned itself, when it has done so at all, with the multiple regression model. Thus, in this proposal, the focus will primarily be upon the creation of a selection battery via the multiple regression model as well, in order to parallel current nonprofessional selection as a first step in examining that process.

Validity of the Selection Battery

There are two ways of considering the validity of a series of selection measures once they have been decided upon. Classical validity examines the correlation of the predicted criterion score with the observed criterion scores--$r_{yy'}$. There is, however, an area which this coefficient doesn't tap--the degree of overlap in prediction which is present between the two populations to be distinguished, i.e., the actual positives and the actual negatives. In any prediction scheme, there is overlap where the people will be predicted to be successful but will be failures and vice versa. Thus, what Wiggins termed "discriminative efficiency" (p. 243) should be examined as well as classical validity. Discriminative efficiency measures the extent to which a set of measures is able to correctly separate and categorize outcomes.

Several options exist when a researcher is attempting to determine the validity of a given group of measures for selection and prediction (Bass, 1971). In predictive validation, the measures to be examined would be administered to a group of applicants prior to selection. The selection process would then take place as is normally done, without regard for scores on the initial measures. Finally, after a set period of time on the job, the selectees' scores would be compared with some criterion measure of their job success and correlations would be determined. In concurrent validation, test scores would be obtained on current employees for whom criterion performance measures are immediately

available. Synthetic validation is the process of inferring validity "in a specific situation from a logical analysis of jobs into their elements, a determination of test validity for these elements and a combination of elemental validities into a whole" (Balma, 1959, p. 395). And, of course, face validity refers to the degree to which a battery appears to be related to outcomes.

All of these types of validation processes have faults. Face validation is often a "hunch" process, and may vary from one researcher to another. The difficulties with predictive validation come with the time lapse involved in waiting for the selectees to have completed a sufficient time in order to obtain accurate criterion scores. However, in concurrent validation, which circumvents this waiting time, there are more serious questions concerning the impact that job experience has upon the scores, as well as the impact of possibly different motivation of the current employees when compared to applicants. Although synthetic validity also requires no great time lapse as well as not needing a large number of workers per position, the validity of the whole may not be exactly equal to the sum of the validities of the parts of a job.

Thus, due to the possible confounding of current job skills and experience with measures obtained in concurrent validation, the quite likely occurrence that more enters into a job than merely the pieces of that job as examined by synthetic validation, and the often nonscientific determination of face validity, predictive validation appears to be the soundest type of validation available

for nonprofessional selection batteries. It would appear that most nonprofessional programs have at best used face validity to justify their selection method; however, in an attempt to quantify the process further this proposal will concentrate primarily upon predictive validity of the measure and/or batteries it creates.

Other Considerations

In addition to examining the validity coefficients of batteries, one must also consider the population to be predicted from and the treatment situation to be predicted to in order to determine if the battery is producing a worthwhile increase over chance selection. First, one must examine the selection ratio, that is, the ratio of applicants chosen for the treatment to the total number of applicants considered suitable. As Taylor and Russell (1939) showed, the usefulness of tests of low validity increases with lower selection ratios. Thus, according to Bass, the lower the selection ratio, other things being equal, the greater the proportion of satisfactory selectees in general, and the more useful even a low validity selection battery may be.

Base rate is another important consideration when deciding on whether to use a particular group of selection measures. The base rate is the proportion of selectees currently in a treatment who are considered to be performing satisfactorily. If nearly everyone is doing well, selection by chance is probably as good a measure as is needed. If almost no one is succeeding, very few

will succeed, and a selection battery of low validity will probably offer little. Therefore, the more moderate (intermediate) the base rate is, the more useful will be a low validity selection battery.

Criterion variability is closely related to base rate. It refers to individual differences in performance in a treatment at present. The greatest potential usefulness of a low validity combination of measures comes with high criterion variability--i.e., when there is "considerable difference in organizational value between the most and least effective person" (Bass, p. 315).

In ultimately deciding on a selection battery, an organization must always keep in mind the cost factor. Is it more expensive to administer a series of tests than it is to risk selecting by chance? According to Cronbach (1970), "Tests with validities from .3 to .5 make a considerable contribution to the efficiency of an institution. The cost savings may be considerable" (p. 429). This is not to purport that all test batteries of low validity should be used. Rather, the researcher must carefully examine all of the above areas in order to determine whether mitigating circumstances will permit the use of measures with low validity. Low validity should not be what the researcher is striving for, but rather an exception to the rule which may not be totally worthless.

Summary

In nonprofessional programs, selections are usually made with a single-stage decision, and when a selection battery is employed, it has most likely been created by the multiple-regression method. Validity of any battery is normally that of face validity, but ideally would be predictive validity. One phase of this research will be the search for a battery for selection of nonprofessionals which yields (1) the highest possible correlation between predicted and actual criterion scores (preditive validity) with the lowest possible intercorrelations among measures in the battery and (2) a significantly greater number of valid positives and valid negatives than false positives and false negatives (discriminative efficiency).

Decision Theory

In the preceding section, types of selection decisions were delineated, methods of creating a battery were discussed, types of validation of such batteries were described and other factors to be examined prior to adopting or rejecting a selection battery were mentioned. In this section, a method of quantifying the usefulness, or cost, of a selection battery thus created to an institution will be discussed.

Once a program director has decided upon an assessment measure or battery of measures, there are still other issues to be resolved. The decision maker must look at outcome possibilities and their meanings or values to the institution involved. The

role of the decision maker then becomes one of assessing the institutional loss and gain as a consequence of adopting a measure. Classical test theory, according to Wiggins, "does not give sufficient emphasis to the fact that the multitude of assignments that are made in any decision situation are seldom of equal importance (value)" (p. 225). Thus, Wiggins broke down the total decision process into five components: information (data available prior to testing as well as the assessment data), strategy (an explicit rule for making decisions), decision (the course of action which the strategy demands), outcome (situation which occurs when certain situations result from the decisions) and utility value (the relative importance of an outcome(s) to an institution). Thus, an institution seeks to adopt the strategy which maximizes the probability of favorable outcomes.

According to Cronbach and Gleser (1965), "The assignment of values to outcomes is the Achilles' heel of decision theory . . . The evaluation of outcomes . . . seems often to be arbitrary and subjective, leading one to question whether any of the conclusions from decision theory can be trustworthy if the starting point itself is open to dispute" (p. 121). However, as Cronbach and Gleser went on to point out, any method of arriving at decisions involves subjective evaluation of some sort. What makes decision theory desirable is that the utilities of different outcomes are explicitly delineated and therefore open to criticism. This fact, however, is much more a credit of the theory than a defect,

for values can be examined and compared much more easily than in those situations where the values are hidden and often unacknowledged.

Wiggins stated that there is a need to first be able to specify precisely the utilities being considered and then to express these utilities in equivalent units in order to permit comparisons. There are two basic ways of determining the utilities of an outcome--either ask the individual what monetary value s/he would place on the outcome or attempt to infer values from the actual decision. Neither approach, though, has supplied satisfactory results as yet.

Wiggins then delineated the appropriate strategy one would use as that which maximizes the function $\Sigma U_i \cdot P(O_i)$, where $P(O_i)$ equals the probability of outcome $\underline{i}$ as determined by base and selection ratios. This function alone, however, does not take into effect the value of the testing as such. In order to take this into account, he proposed the following equation:

$$EU = U_1 \cdot P(VP) + U_2 \cdot P(FP) + U_3 \cdot P(FN) + U_4 \cdot P(VN) - U_t$$

or "The expected utility of a decision strategy is equal to the sum of the products of the probability of each outcome and its associated utility, minus the cost of testing" (p. 258). Thus, if a decision maker has several testing procedures, s/he is able to compare the expected utility of each strategy (once s/he has

assigned utilities to each otucome and determined the probability of each outcome--by no means a small task).

Summary

Thus, in these sections we have developed a sequence which a program director would ideally follow prior to deciding upon a method of selection. S/he would administered measures to all applicants, select as is usually done (or, ideally, select all applicants), and examine the relationships of the measures to outcome (success on the job) as well as the ability of the measures to discriminate between the actual successes and actual failures. Then, once the decision maker has carefully created and validated her/his selection measures, s/he would examine the base and selection rates, estimate the probabilities of the possible outcomes, determine the utilities of each outcome to the organization, determine the utility of the assessment strategy and compute the expected utility of each decision strategy open to her/him. With this evidence before her/him, s/he should then be able to chose that strategy which maximizes the favorable outcomes for the institution.

The preceding discussion has been very general. In the following section, the selection procedures for programs using nonprofessionals will be examined, weighing advantages and disadvantages of the prevalent selection methods.

Nonprofessional Selection

In the above sections, general principles of selection and decision theory were delineated. Most nonprofessionals programs make single-stage decisions; however, their methods and strategies differ. This section will survey modes of selection common in nonprofessional programs.

Selection of nonprofessionals seems to be one of the things that everyone talks about but no one does anything about. Many authors have acknowledged the importance of the selection proceedings. For example, Johnson (1971) stated, "The selection of persons to be trained must be done with extreme care" (p. 234). Delworth and Moore (1974) claimed, "Accurate selection of student trainees is the key to an effective training program" (p. 429). Aiken, Brownell and Iscoe (1974) purported that "the success of a training program starts with the selection of trainees" (p. 481). Carkhuff (1969) noted,

> Unless we make the assumption that all persons are capable of being trained and, ultimately, of functioning effectively in the helping role--anad the increasing number of persons who need help makes this assumption a difficult one--we must develop selection indexes that are relevant, meaningful and valid for purposes of helping. (p. 79)

Even when the author was not especially concerned with choosing the suitable trainees for her/his training program, s/he often noted the importance of choosing appropriate people to work in the program. Goodman (1972) remarked,

> Developing reliable selection procedures becomes especially important where training is minimal. Indeed, training may be less important than appropriate selection in programs

> using nonprofessionals--particularly since therapeutic competence probably is more a matter of untutored talent or interpersonal sensitivity than of skill learned during a training course. (p. 27)

Dorr, Cowen, Sandler and Pratt (1973) went one step further in urging researchers to not merely be concerned in selecting appropriate trainees for their training programs, but to connect nonprofessionals' attributes with outcomes in order to determine ultimate success. Durlak (1971) offered the situation as a challenge:

> One of the most pressing and at the same time difficult problems for investigators to overcome is that of developing adequate selection procedures for choosing nonprofessionals to serve in the variety of programs which could make use of such personnel. It is especially difficult since there are few guidelines to be found in professional literature. (p. 229)

It would seem that when a program director is considering a selection technique, there would be several issues of importance s/he should consider. First a useful selection technique would ideally be able to differentiate potential "dropouts" from successful completers--those who, if they begin the training or service program, will eventually terminate themselves. Another useful type for a technique to point out would be the "compliant" type--which applicants will adapt themselves most easily and effectively to the program being offered. Still another crucial consideration of a selection technique is, of course, that of "outcome"--how well does the technique predict who will have success with the target population. Finally, of course, one must weigh

the cost of the strategy against the benefit of the prediction of the strategy.

Types of Selection Prevalent in Nonprofessional Literature

Unfortunately, literature concerning nonprofessional programs is not especially edifying when studying the selection of people to work in such programs. As Goodman (1967) stated, "We know little about how to study nonprofessional programs and less about their effectiveness and most about their adventures in getting off the ground" (p. 1772). In general, "The trend has been toward descriptive articles which may make practical recommendations to those selecting and training preprofessionals for human services" (Codori & Cowles, 1971, p. 48). Overall, there seem to be three major types of selection techniques utilized when choosing nonprofessionals: self-selection, interview and/or recommendations, and psycholgoical testing. The next sections will examine these methods and delineate major advantages and disadvantages of each.

Self-Selection. Table 1 shows that 24 of the 40 studies examined either did not even specify their method of selection or used self-selection as their screening procedure. Especially the earlier programs often incorporated any students who were available and interested (Holzberg, Knapp, & Turner, 1967; Umbarger, Dalsimer, Morrison & Breggin, 1962). In these instances, the program directors believed that the student had shown sufficient motivation by coming to the institution involved to make her/him an impactful change agent. This line of thinking seems to be

Table 1

Studies Using Self-Selection (or No Selection Reported)

Author	N	Target population	Findings
Aiken, et al. (1974)	10	college students	*
Brown (1965, 1974)	6	college students	Freshmen counseled be peers gained in attitude and study skills.
Brown, et al. (1971)	*	college students	Paraprofessionals are effective.
Buckley, et al. (1970)	*	chronic schizophrenics	Patients improved.
Chinsky (1968)	30	chronic hospitalized mental patients	GAIT "warmth" and "understanding" were positively related to patient improvement
Dawson (1973)	20	college students	*
Holzberg, et al. (1967)	*	chronic mental in-patients	Patients and students changed.
Kalafat & Tyler (1973)	*	college students	*
King & Turner (1975)	1	mentally retarded adult	Social reinforcement can be effective with a profoundly retarded adult.
Knapp & Holzberg (1964)	85	chronic mental patients	Volunteers were not significantly different from nonvolunteers.
Kopita (1974)	42	college students	Those students trained changed more than those not trained.
Kreitzer (1969)	21	hospitalized children	Nonprofessionals worked with behavior therapy.

*not explicity specified

Table 1 (Cont.)

Author	N	Target population	Findings
Lawton & Lipton (1963)	6	mental inpatients	*
Levine (1966)	*	mental inpatients	*
Newton (1974)	32	college students	Training increased level of communication.
Poser (1966)	11	male chronic schizophrenics	Nonprofessionals had more success than professionals.
Scheibe (1965)	99	chronic mental inpatients	College students crystallized their goals as a result of their experience.
Schnelle, et al. (1975)	4	mental health center clients	College students can be effectve as behavioral engineers.
Spoerl (1968)	25	mental patients	All felt the experience was beneficial to all.
Umbarger, et al. (1962)	> 2,000	mentally ill hospitalized adults	*
Vesprani (1969)	33	female mental inpatients	Nonprofessionals learned accurate empathy.
Walker, et al. (1967)	*	mental inpatients	*
Wasserman, et al. (1975)	*	college students	*
Zunker & Brown (1966)	8	college students	Paraprofessional counselors were as effective as professional counselors.

*not explicity specified

linked to those who believe that nonprofessionals, especially college students, have as their greatest asset a fresh approach and more "earthy" qualities bred or trained out of professionals.

The major advantages of this method of selection are its ease of "administration," low cost and lack of administrative entanglements surrounding the rejection of an applicant. Disadvantages of self-selection are the increased chance of false positives in selection, and the lack of empirical evidence pertaining to accurate prediction of successful outcome.

Selection by Interview and/or Recommendations. Of the surveyed programs, approximately one-third (13 of 40) specified selection by interview alone, recommendations alone, or a combination of the two (e.g., Cowen, Zax, & Laird, 1966; Klein & Zax, 1965; Persons, Clark, Persons, Kadish, & Patterson, 1973). These researchers have "relied upon clinical judgment gleaned from interviews designed to evaluate broad personal qualities such as motivation, interest and ability to work with others" (Durlak, p. 229). Thus, these researchers attempt to screen for gross maladjustment or gross unsuitability for the program they are running.

Advantages of this approach to selection are primarily the ability to screen out likely false positives (those seemingly unfit for service) and the ease of application. The major disadvantages are the high actual cost of professional time spent in interviewing, the lack of specific objective criteria for selection and the lack of empirical evidence of accurate prediction of success.

Table 2

Studies Using Selection by Interview and/or Recommendations

Author	N	Target population	Findings
Bohr (1972)	*	adolescents	Students learned and adapted to their positions well
Cowen, et al. (1966)	17	emotionally disturbed children	Talking is correlated with child improvement.
Davison (1965)	4	autistic children	Students learned behavior modification techniques.
Goodman (1967)	99	"troubled" boys	GAIT scores are a weak predictor of success with boys.
Huer (1973)	16	college students	*
Keeley, et al. (1973)	37	behaviorally disordered children	*
Klein & Zax (1965)	13	mental inpatients	Students learned.
Mitchell (1966)	*	troubled children	*
Persons, et al. (1973)	8	college students	Service at the counseling center increated and clients liked the nonprofessionals.
Rappaport, et al. (1971)	30	chronic mental patients	College students had impact.
Stuckey, et al. (1971)	20	psychotic children	*

*not explicity specified

Table 2 (Cont.)

Author	N	Target population	Findings
Suinn (1974a)	20	pre-school and kindergarten children	Most problem behavior improved.
Zax & Cowen (1967)	17	emotionally disturbed primary grade children	The college students changed during the program.

*not explicitly specified

Selection by Psychological Testing. Table 3 shows only three of the 40 reviewed selection procedures used some kind of psychological testing of the applicants. Roberts (1976) initially tested undergraduate dormitory resident assistants (RA's) on communication level, physical condition, grade point average during the preceding three terms, and a measure of "cooperativeness" during the selection process. However, due to lack of variance on the first two variables, she selected the RA's on the basis of mean grade point average and "cooperativeness" alone. She assigned RA's with high levels of each criteria to an experimental condition, and RA's without both criteria to the control condition. The experimental group received 16 hours of communication skills training and four hours of referral skills, decision-making and program development skills. The control condition received no training. Using job performance as outcome, she found that the experimental RA's had higher assessed job performance after training than did the control RA's after no training. Needless to say, the non-random assignment to conditions clearly muddles the outcome of the effect of the selection measures.

Stollak (1968) appears to have conducted one of the few studies which overtly selected the participants by testing. He had 75 undergraduate volunteers interested in learning play therapy techniques. All 75 took Smith's Sensitivity-to-People test; those five males and five females with the highest scores on this measure were designated High Potential Therapists (HPT); the five male and five females scoring the lowest were designated Low Potential

Table 3

Studies Using Selection by Psychological Testing

Author	N	Target population	Findings
Roberts (1976)	20	college students	Subjects with higher grade point averages & higher "cooperativeness" changed more after training than did subjects with lower grade point averages & lower "cooperativeness" subjects without training.
Stollak (1969)	18	psychological clinic-referred children	Undergraduates learned play therapy techniques.
Stollak (1975)	20	psychological clinic-referred children	Trainees exhibited more acceptance, self-direction and involvement with the children than did untrained controls.

Therapists (LPT). Two thus selected dropped out at this point; two others dropped later; four student cases were terminated by the parents. Thus, the study encompassed five HPT's and 7 LPT's. Stollak found there was no difference in performance of the HPT's as compared to LPT's after training. Thus, the selection device chosen by Stollak apparently did not successfully discriminate between effective and ineffective play therapists.

In a later study, Stollak, Scholom, Green, Schreiber, & Messe (1975) selected participants using a Parent Attitude - Research Instrument, a Sensitivity to Children projective questionnaire and a Personality Questionnaire to assess general mental health. Selection occurred in a similar manner: the 10 males and 10 females with the highest scores on all measures were designated High Potential (HP); the 10 males and 10 females with the lowest scores were designated Low Potential (LP). Ten from both categories were randomly selected (controlling for equal numbers of male and female) as trained experimentals; the remaining ten of each category served as untrained controls. Once again, though, HP's and LP's were not significantly different as differentiated by the selection battery.

Although Suinn (1974a, 1974b) selected his participants by rating interview only (see Table 2), he did gather data on all applicants in order to examine possible relationships. He administered a Dogmatism scale, the Edwards Personal Preference Schedule, attitude toward elementary school teachers scale and work samples during training (for those selected). Those accepted for

training were categorized by their job performances into one of three levels: high (qualified as a consultant), moderate (some minor problems) or low (not qualified as a consultant). Using the multiple regression method, Suinn found that when agency staff rated job performance, the single most powerful predictor was the Dominance score on the Edwards Personal Preference Schedule (low scores predicting high job rating; multiple r = .85). When the training staff rated job performance, however, a work sample (appropriate interviewing) was the most powerful predictor (multiple r = .91). Again, since trainees were not randomly selected, it is difficult to make any sound generalizations from this study.

As with Suinn, Rappaport, Chisky, and Cown (1971) did not screen their applicants using test scores, but rather interviewed them to eliminate those obviously unfit. However, they were interested in prediction, so they also obtained various measures upon their selectees. One area of primary interest was that of assessing "therapeutic talent." Rogers (1961) and Truax and Carkhuff (1967) showed that therapeutic talent was mainly based upon three basic attributes: understanding, acceptance and genuineness. Thus, the researchers employed a Truax scale of dimensions of accurate empathy, nonpossessive warmth and congruence in order to attempt to quantify these characteristics in their workers. Also, they obtained scores on the Whitehorn and Betz A-B scale, measuring interaction style; Jackson's Personality Research Form; Rotter's Internal-External Locus of Control scale; Chapin's Social Insight Test; Jourard's Self-Disclosure Inventory; Wrightsmann's

Philosophies of Human Nature Scale; an adjective check list; and a semantic differential attitude scale. After factor analysis of the above measures to produce six main factor scores, they were correlated with patient change scores. Although no statistics were reported, the authors stated "The number of significant correlations between these factor scores and patient change scores on the eleven test performance and five ward-behavior measures of chronic schizophrenic patients failed to exceed chance expectancy" (p. 122).

Overall, the major advantages of selection by psychological testing seem to be the appearance of empiricism, the ease of administration of paper and pencil measures and the lack of expense involved. The disadvantages, however, are the amount of time spent by the respondent in filling out measures and, as shown clearly above, the lack of discriminative power and prediction ability.

Selection by Behavioral Measure. Obviously, the three prevalent methods of selecting nonprofessional for participating in human service positions are fraught with pitfalls, with the largest pitfall being that none of them predict successfully to outcome. In response to this problem, Goodman (1967) decided to venture into a new area for selection--that of a behavioral measure. As did Rappaport, et al., he concerned himself with the three Rogerian attributes deemed necessary for therapeutic talent--understanding, acceptance and genuineness. In order to attempt to

quantify these traits in nonprofessionals, he created the Group Assessment of Interpersonal Talent (GAIT). Goodman (1967) described his measure as follows:

> Each session is attended by eight applicants and three staff members. Each member is given a card and asked to fill it out with a description of one of his interpersonal concerns that can eventually be read to the group . . . An applicant, chosen at random, is asked to read his card to the group. We call him the "discloser." Any other applicant can make an effort to understand how the discloser feels about the problem read to the group by engaging him in a five-minute dialogue. We call the second person the "understander." Understanders are instructed not to give advice or interpret and to avoid asking many questions. Thus, one person is attempting to solve the problem of how to disclose or to be genuine in a manufactured group situation, while the other is attempting to solve the problem of how to listen . . . The procedure continues around the group . . . and everyone tried each task once. At the end, all applicants rate each other, and the three staff members rate the applicants. Ratings are done on a sociometric type instrument, with items such as "He really seemed to understand what the other person meant" with six-point scales ranging from "much like him" to "not like him." Items cover the areas of warmth, self-disclosure,empathy, rigidity, surgency, and so on. (p. 1773-1774)

Goodman (1972) has stated, "Most of the findings on the GAIT suggest a coherent internal order among the GAIT items . . . Our study of the relation to the indices offers some encouragement about the validity of the GAIT. The reliability findings, the meaningful intercorrelation matrix, the consistency of the small test-retest sample, the significant correlations between staff and student ratings have also provided clues to the GAIT's promise as a worthwhile new measure" (p. 40).

Despite Goodman's glowing comments about the GAIT, however, it has run into three severe problems in proving itself a "worthwhile new measure." First, interrater reliabilities were found

to be "spotty" by Rappaport, et al.: "A few were sufficiently low to raise serious question about the scale's usefulness" (p. 118). For example, interobserver reliabiltiy of Open was .34, of Relaxed, .19. When observer and peer ratings were averaged, the highest interrater correlation was .59 (Meaningful), with other scales ranging from .30 (Best Counselor) to .53 (Warm). D'Augelli, Chinsky, and Getter (1974), in response to this criticism, stated that "attempts are being made to refine the GAIT as an assessment procedure with specific emphasis on increasing the objectivity of the rating system" (p. 64).

A second criticism comes from Dooley (1975). Although Goodman mentions small sample test-retest calculation (N = 41, three-week interval), Dooley found that test-retest correlations on peer ratings overall varied widely when done on a nine-week interval. Correlations ranged from a quite reliable Quiet (.79) to Blue (.08), with one scale having a significant correlation at $p = .05$ level (Rigid--.38) and three at $p = .01$ level (Relaxed--.48; Meaningful--.55; and Quiet--.79). Since Dooley found significant correlations over a three-week interval on all scales, he claimed that the higher test-retest correlations at the shorter interval may point to the fact that the raters were attempting to "remember" what they had rated the person earlier and to appear consistent. However, it is also possible that the traits measured by the GAIT are not stable as measured.

The third, and most serious, problem with the GAIT is its lack of predictive power. Goodman (1972) looked at companionship

therapy conducted by college males with troubled boys, referred by schools. He used as criterion measures in his project

> (1) two retrospective change scores of self-esteem . . . based on observations of teachers and parents; (2) a pre-post discrepancy measure on the adjustment scale of the Adjective Check List (ACL) taken by the parents; (3) a composite school-aggression score built from classmate and teacher response, pre to post, on the Peer Nominations Inventory . . . (p. 44)

However, in examining correlations of the GAIT ratings with these outcome measures, Goodman found only nine of 48 correlations significant: for observer ratings,

GAIT Open	-- ACL adjustment scale	(.26)
	-- Composite school-aggression score	(-.31)
GAIT Understanding	-- ACL adjustment scale	(.20)
GAIT Therapeutic Talent	-- ACL adjustment scale	(.26)

and for combined observer/peer ratings,

GAIT Open	-- Composite school-aggression score	(-.24)
GAIT Understanding	-- + change in self-esteem	(.20)
GAIT Therapeutic Talent	-- + change in self-esteem	(.20)
	-- ACL adjustment scale	(.24)
	-- Composite school-aggression score	(-.20)

Along the same lines, Rappaport, et al., attempted to relate personality characteristics of their college student volunteers in a state mental hospital to their success as group leaders of their patients. Two types of outcome data were used. First, a mean group improvement was obtained by subtracting the pretest mean group score on a series of measures (e.g., reaction time, tapping speed, perceptual and perceptual-motor tests and verbal fluency) from the posttest mean group scores. The second type of

data was ratings of ward behavior (mood, cooperativeness, communication, social contact and total adjustment). Although no specific correlations were reported for the first outcome measure, alledgedly only eight of the 110 correlations between each GAIT MEASURE (combined observer and peer ratings) and mean group improvement scores were significant. Similar lack of results occurred with peer and observer ratings alone. In addition, only one of 50 correlations between combined peer and observer ratings and the five ward behavior ratings were significant; similar results were obtained again from peer and observer ratings alone.

When looking at mean percentage improvement scores on ward behavior measures, Rappaport, et al., reported 16 significant correlations in the 150 correlations between observers, peers and combined observers and peers GAIT ratings and ward behavior. The significant correlations for observer ratings were:

GAIT Understanding	-- Mood	(.48)
GAIT Warm	-- Mood	(.39)
	-- Cooperation	(.41)
	-- Total adjustment	(.46)
GAIT Best Counselor	-- Mood	(.40)
	-- Total adjustment	(.42)
GAIT Therapeutic Talent	-- Mood	(.37)

For peer ratings alone, the significant correlations were

GAIT Quiet	-- Cooperation	(.37)
GAIT Rigid	-- Social contact	(-.57)

For combined ratings, the significant correlations were:

GAIT Understanding	-- Mood	(.36)
GAIT Quite	-- Cooperation	(.37)
GAIT Warm	-- Mood	(.38)
	-- Total adjustment	(.37)
GAIT Rigid	-- Social contact	(-.49)

GAIT Best Counselor	-- Mood	(.43)
GAIT Therapeutic Talent	-- Mood	(.37)

Using a shortened version of the GAIT rating scales, Chinsky and Rappaport (1971) found with the same target population and nonprofessional population only four out of 15 significant correlations of observer ratings to ward behavior change ratings:

GAIT Understanding	-- Mood	(.48)
GAIT Accepting-warm	-- Mood	(.39)
	-- Cooperation	(.41)
	-- Total adjustment	(.46)

Peer ratings were found to have no significant correlations with outcome, and neither observer nor peer ratings had any significant correlations with patient change in test performance (mean group improvement score described above).

Goodman himself (1972) stated, "The GAIT was a late arrival; most of our research on its measurement properties has a patchwork quality. A comparison of traits between different methods could reveal patterns of convergence and discrimination yielding powerful evidence for validity" (p. 38). However, this evidence has not been forthcoming; on the contrary, researchers who are investigating the measure seem to be uncovering contradictory evidence.

Summary

Thus, there are three prevalent methods of selection of nonprofessionals: self-selection, interview and/or recommendations, and psychological testing. None of these has proven to be significantly predictive to outcome. In an attempt to overcome

this difficulty, Goodman created a behavioral measure--the GAIT--which seemed to hold promise; however, it too has run into similar problems--low interrater reliability, low test-retest reliability, and no consistent significant prediction to outcome. Onge again, the nonprofessional selection field finds itself without sound prediction and decision procedures.

The Need for a Change

As has been shown in the previous literature review, there is a need, especially in programs utilizing nonprofessionals, for new selection measures. Mordock and Platt (1969) stated "Early in our research, we found that instruments to measure personality are of little value in predicting success or failure" (p. 228). Rappaport (1977) concurred that it has been found that "personality testing did not predict success. Individual interviews, although probably better than personality measures, have also not been shown to be either reliable or good enough predictors to make them worth the relatively high cost of professional time" (p. 385). Zax and Specter (1974) claimed program directors who have created selection strategies have done so without sound research--that they have begun "with preconceptions of qualities that they feel will be desirable and set up elaborate procedures to select for them" (p. 381). Holzberg, et al., stated "Criteria for effective screening are frankly unknown. While our research may ultimately yield predictors of success, neither the predictor nor the criteria have as yet been delineated" (p. 93).

The GAIT was an attempt to address this issue; however, as was shown above, it too has fallen far short of its mark. In addition, Goodman's assumptions of the necessity of Rogerian attributes for a relationship are not the assumptions of the Adolescent Diversion Project. While these attributes may be important, the Project was also interested in quantifying other areas of interactions, such as questioning style, planfulness and resourcefulness, which are not addressed in the GAIT.

However, it would appear that Goodman was on the right track in creating a behavioral measure. Rotter (1960) put forth the notion that prediction from testing would be most effective when the testing situation is similar to the criterion stiuation. Mischel (1968) concurred: "Predictions tend to be best when the predictor behavior is sampled in situations that approximate the criterion situation as much as possible" (p. 278). Durlak also purported that the best way to predict future behavior in a particular situation is to obtain a measure of present behavior in the same or similar situation: "a behavioristic measure of the trainee functioning in a simulated or real-life situation may be the most effective way of selecting potential therapeutic personnel" (p. 237). According to Guilford (1959), the basic assumption in using a behavioral measure (situation test) is that the researcher needs to see the person performing in a complex, lifelike situation in order to obtain a more realistic view of her/his way of relating to others. Nunnally (1970) warned that observations in a contrived situation are not meant to measure personality traits, but rather

to measure traits relating to particular situations. Due to the lack of predictive power in personality traits as shown above, though, this appears to be a fair and acceptable condition. Thus, while Goodman may have created a situation close to the criterion performance of his volunteers, the situation is not close enough to the criterion performance of the nonprofessionals in the Adolescent Diversion Project to allow a measurement of all of the categories of behavior of interest.

One major criticism of a behavioral measure is its expense. Admittedly, a behavioral measure can cost more to administer than a paper-and-pencil test. However, Wiggins has shown that a more expensive, but more valid, measure can actually have a higher utility in the overview of the decision strategy. Thus, if one is able to create a measure which predicts more consistently to outcome and/or discriminates successes from failures more accurately, it is possible that money and time could be saved, when viewed from the utility to the Project.

Therefore, this research proposed two facets: one, the development of a behavioral measure more applicable to the needs of the Adolescent Division Project, and two, hypothesis-testing, comparing the created measure and other more traditional methods of assessing nonprofessionals with outcome measures of the adolescents with whom they were working.

Specifically, first a personality measure--the Personality Research Form (Jackson, 1974)--was examined. This is a

questionnaire developed by the rational-empirical technique (Jackson, 1970), which generates 14 factors of personaltiy: e.g., achievement, affiliation, autonomy. This measure was administered approximately three months prior to assignment to condition. Secondly, a behavioral measure was created to examine the undergraduates' interpersonal styles while interacting with two adolescents. From this measure, components were generated for the purpose of comparison. This measure was administered approximately eight weeks after assignment to condition.

Thirdly, peer, supervisor and self evaluations of students' skills while working with their assigned youths were examined. These measures were chosen to gauge how those working closely with the interveners viewed their skills as related to the rest of the supervision group. Peer, self and supervisor in-class evaluations were gathered at three intervals during the three-term commitment--immediately after training, at a midpoint between the end of training and completion, and at completion of the three-term commitment. Fourthly, process measures were obtained concerning the intervention from interviews conducted with the assigned yough, her/his parent, a peer nominated by the youth and the college student assigned to the youth. With these measures, it was possible to examine how the student was getting along with the youth and what was being done as perceived by those four sources. These interviews were conducted at three points: six weeks after intervention began, 12 weeks after intervention began and at termination (18 weeks after intervention began).

The outcome criteria focused on the assigned youth. Data were gathered on their police contacts while involved with the college student (number and seriousness of contact), their court contacts while involved with the college student (number and seriousness of contact), and their school records (grade point average, percentage of days attended school and number of credits earned toward graduation). This information was then used to classify each case as either a "success" or a "failure." Thus, rather than merely creating an analog by examining simply how the measures on the college students intercorrelated, it was possible to measure actual impact of college students on the target population.

Obviously, the first four measures spanned approximately 12 months. In this way, there was a chance to examine the students at different points in their involvement with the Project, and therefore the options of looking at any of these points in relation to outcome criteria as well as relations among measures. Specifically, the relationship between the Personality Research Form and the factors of the behavioral measure was investigated. Similarly, the relationships between the behavioral measure and other measures (peer, supervisor and self overall in-class evaluations and process interviews) were scrutinized separately. In addition, the relationships of all the measures with the outcome measure were examined, using step-wise discriminant function analyses in order to attempt to create a selection battery as suggested in the previous sections.

CHAPTER II

METHOD

Setting

The Adolescent Diversion Project was an attempt to divert adolescents from the juvenile justice system via intervention on a one-to-one basis of an undergraduate for 18 weeks. Upon successful completion of the 18-week program, the adolescent's record with the justice system was expunged.

The Project was staffed by the project director and seven graduate students who served as research coordinators and supervision coordinators in the educational pyramid paradigm, as set forth by Seidman and Rappaport (1974). In addition to examining different intervention effects with adolescents, the Project also examined different intervention techniques and training techniques. For example, during the proposed research period, the Project examined the effect of low intensity training (three weeks of training, supervision meetings once a month) versus high intensity training (eight weeks of training, supervision meetings once a week). Within the low intensity condition, the Project researched the effects of small group supervision meetings (7-8 people per group, 2 supervisors) versus large group supervision (15 people per group, 2 supervisors). Within the high intensity condition, the Project also researched the effects of an action-oriented

intervention (behavioral contracting and child advocacy) versus a relationship-bulding intervention (active listening and problem-solving skills). Thus, there were four different training conditions: Low Intensity--Large Group, Low Intensity--Small Group, high intensity--action (Action), and high intensity--relationship (Relationship).

In addition to looking at outcome, the Project also examined the process of the intervention by way of interviews conducted at four time periods (beginning, six weeks, 12 weeks and termination) with the assigned youth, a peer nominated by her/him and the parent of the youth assigned; and at three time periods with the assigned college student (six weeks, 12 weeks and termination). Each interview was conducted informally, tape recorded and coded by a trained undergraduate interviewer. One interviewer was assigned to conduct all interviews for each case. Areas explored in the interviews were family, school, self-report delinquency, and the intervention process, among others.

Subjects

The subjects were undergraduates at Michigan State University who were first enrolled in Psychology 370, Section 1, during Fall term, 1977. Figure 1 presents a flowchart of the selection process for the undergraduates. Students were recruited by the mailing of a new course announcement to all Freshmen, Sophomore and Junior social science majors at Michigan State University during May 1977. Approximately 450 students responded initially to the

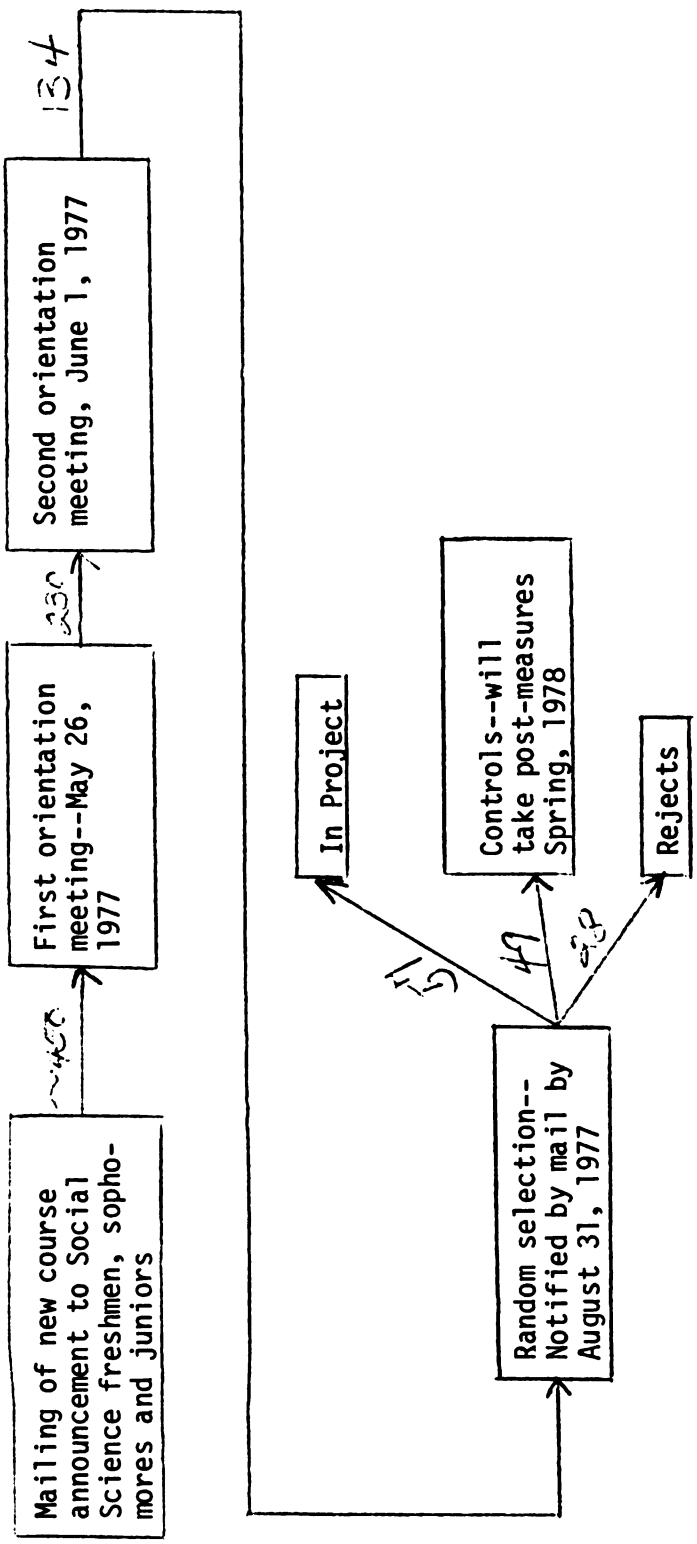


Figure 1. Flowchart of Undergraduate Selection.

letter by phone, expressing some interest in participating. Interested students were subsequently required to attend two orientation and assessment sessions. Approximately 300 students appeared at the first general orientation meeting on May 26, 1977. At this meeting, the project director spoke to the students regarding the background of the Project and requirments of participation in the Project; later, students' questions were answered. Finally, demographic data and career goals were collected, people signed contracts stating both their willingness to commit themselves to three terms in the course (Fall 1977, Winter 1978 and Spring 1978) and their willingness to participate in further assessment measures. A second meeting time was announced at this time for further measures. Approximately 70 students left the meeting without filling out the initial data; thus, demographic data, career goals and contracts were filled out by 230 students. Of the 230 people who filled out the measures in the first meeting, 134 students returned for the final meeting (June 1, 1977), where various questionnaire measures were administered-the Personality Research form, a version of the Semantic Differential adapted to the area and the target population, a Delinquency Orientation scale attempting to assess the respondent's stance on treatment of delinquents, Rotter's Locus of Control Scale, a portion of the Strong Vocational Interest Blank to measure the respondent's amount of introversion or extraversion.

Obviously, this "recruitment" procedure in itself proved to be quite a selection device. From the "self-selected" pool of 134, 58 students were chosen randomly, controlling for equal numbers of each sex. Shortly after the registration period for Fall term, one female chosen for the Relationship condition decided not to participate. Due to difficulties arising from trying to find another person at that late date, she was not replaced. Thus, Fall term found the Project with 57 students; 13 in the Relationship condition; 14 in the Action condition; 15 in the Low Intensity--Large Group condition; and 15 in the Low Intensity--Small Group condition. By the end of the school year, two more undergraduates dropped out, both due to personal problems developing around the end of Fall term and the beginning of Winter term, 1978. Thus, the final split of students was: 12 in the Relationship condition; 14 in the Action condition; 15 in the Low Intensity--Large Group condition; and 14 in the Low Intensity--Small Group condition.

Students not selected were randomly placed either on a Control list or a Reject list. (Due to the number of female applicants, there were more females than needed to serve as controls; so, in order to save time and money, this distinction was made.)

Also of interest in this research was the manner of selecting the sample of the target population--the juveniles. The adolescents were chosen for the Project in the manner depicted in Figure 2. Adolescents who had pleaded guilty to a petitioned

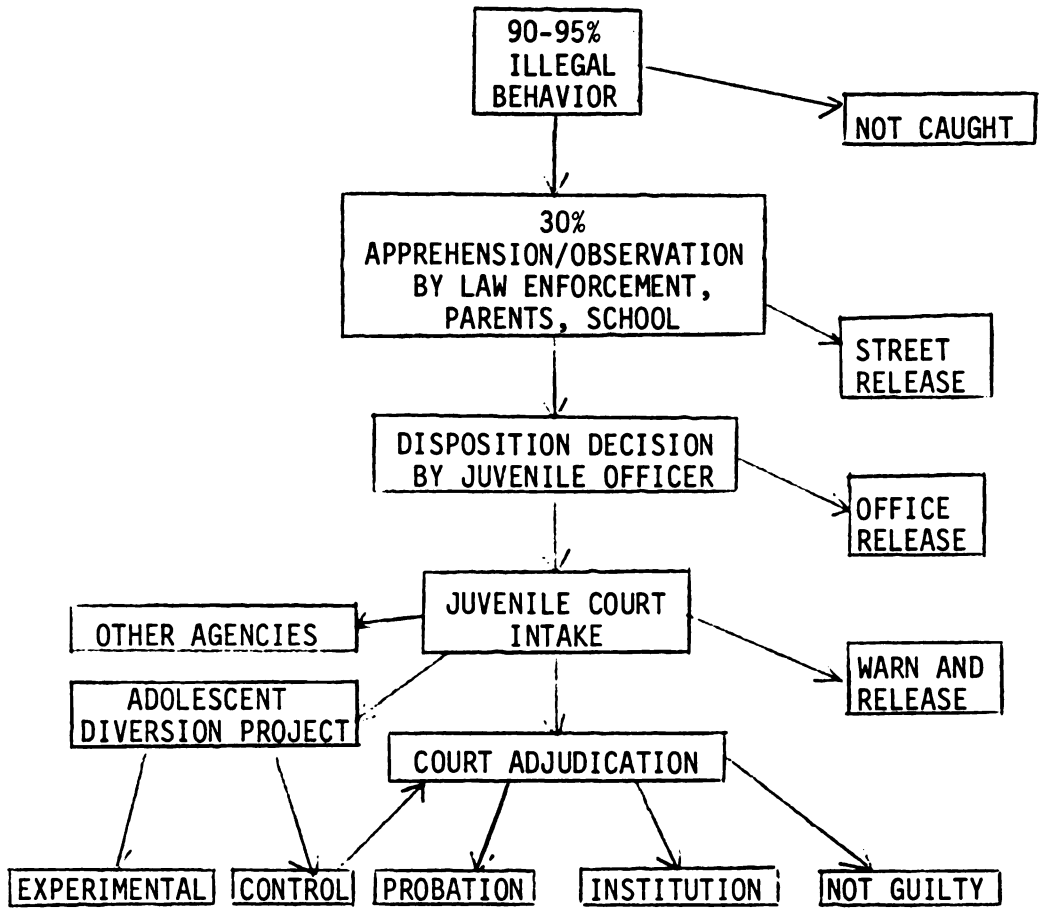


Figure 2. Flowchart of Adolescent Selection.

charge in Ingham County Probate Court and who were deemed appropriate for the Project by the court personnel (e.g., not guilty of extremely minor or major charges) were referred to a Project Intake interviewer. This interviewer, a Project staff member, met with the youth and her/his parents to explain the option of participating in the program. If all involved were interested in being considered for the Project, they signed a contract acknowledging their interest and affirming their knowledge of their rights as participants. They were then assigned randomly to either an Experimental or a Control group by the draw of a sealed envelope, containing a slip saying either "Project" or "Control." If they were selected to be in the Project, they were then further assigned randomly to one fo the four training conditions.

Seventy-seven delinquent youth were referred to the Project from October, 1977 through January, 1978. Four youth declined to participate. Those deciding that they wanted to be part of the Project had the following characteristics: 62 were males and 11 were females; 51 were Caucasian and 22 were non-Caucasian; the average year in school was ninth grade, with a range from fifth to twelfth grade; 65 percent of the youth came from broken homes; and the average age was 14.3 years old. Criminal activities covered a broad range, but tended to be nonserious and serious misdemeanors and nonserious felonies. There were no significant differences among the youth in the experimental groups, or between the experimental and the control groups on any of these demographic variables.

The Adolescent Diversion Project had been reviewed by the Uniersity Committee on Human Subjects in Research and had met the required criteria.

Design

The design of this research had two major components. First, a behavioral measure was developed which rated undergraduates on twenty categories of behavior when interacting with an adolescent. Once the measure was created, the research turned to hypothesis-testing. Scores from the behavioral measure were compared to other measures obtained during the one-year span of the subjects' involvement in the Project: (1) the Personality Research Form (PRF), given prior to assignment to condition; (2) peer, supervisor and self in-class evaluations given three times during the three-term course; (3) the Intervention Scale from process interviews obtained three times during the student's involvement with the youth; and (4) an outcome measure of success or failure of the youth during the intervention period.

One design to be examined was the predictive capabilities of the behavioral measure and the first three measures in relation to the outcome measure. Thus, in this design, the independent variables were scores on the PRF, peer, supervisor and self in-class evaluations, the Intervention Scale from the interviews and the behavioral measure; the dependent measure was outcome.

However, with the above variables, there were other optional designs of interest as well. One involved exploring the

relationship among the behavioral measure, the PRF, peer, supervisor and self in-class evaluations and the process interview data. In addition to other combinations, there was also the issue of the effect of the various training conditions on the behavioral measure--e.g., did the student's training make her/him appear different in the behavioral measure from students in other training conditions? (This was important in this research since, because of the relatively late gathering of the behavioral data, training conditions might have influenced the measure scores.) Thus, in this research, data were examined in numerous ways in an attempt to discover relationships rather than to confirm expected relationships.

Procedures

Recruitment

As mentioned above, students were recruited initially through a mailing of a new course announcement. They were required to attend two meetings and fill out demographic information, career goals, and five psychological questionnaire measures.

Personality Research Form (PRF)

Measure development. Form A of the PRF (Jackson, 1974) is a 300-item test where respondents mark "True" or "False" to statements. The responses generate scale scores on 14 general "traits:" achievement, affiliation, aggression, autonomy, dominance, endurance, exhibitionism, harmavoidance, impulsivity, nurturance, order, playfulness, social recognition, and

understanding plus an infrequency score (adpated from the MMPI Lie scale). Jackson (1974) reported test-retest reliabilities (N = 135, one week time lapse) for a longer version of the PRF (Form AA). Looking only at the scales included in both Form A and AA, reliabilities ranged from .77 (autonomy) to .90 (harmavoidance).

Of all of the personality measures available, this measure was chosen as an independent variable because of its careful creation (Jackson, 1970), its already factored scales, and its applicability to a normal population. Since most non-professional programs which do any selection use personality measures, it was decided to include a personality measure to be compared to the other measures for effective prediction.

The PRF was administered during the second meeting of recruitment, prior to assignment to condition. (See Table 4 for a schedule of measures administered to the undergraduates.)

Development of factor scores. Once the data had been gathered, the raw scores for each scale were transformed into T-scores, using tables compiled and reported in Jackson (1974) derived from male and female norms. These standard scores were examined using a principal components analysis with varimax rotation. Employing the Kaiser criteria (Kaiser and Caffry, 1965), four components were extracted: Achievement-motivation (achievement, endurance, dominance, understanding--loadings ranging from .81 to .58); Impulsivity (impulsivity, nor harm-avoidance, not order, play--loadings ranging from .80 to .49);

Table 4

Undergraduate Measures Schedule

	(6/1/77) Prior to selection	(11/8/77) Six weeks after training begins	(11/28/77) End of training	3/15/78) Midpoint	(6/8/78) End of course
Personality Research Form	X				
Behavioral Measure		X			
Peer and Supervisor Nominations			X	X	X
Outcome of assigned youth					X

Outgoingness (affiliation, not autonomy, social recognition, exhibition--loadings ranging from .78 to .56); and Aggression (aggression, not nurturance, loadings of .79 and -.57). (See Table 5.) For each of these components, an average component score was computed by summing the unit-weighted T-scores of the component and dividing by the number of components loading on that component. Thus, the 14 scales of the PRF were collapsed into four average component scores. Table 6 offers a summary of the component scores; Table 7 exhibits the intercorrelations of the component scores.

Behavioral Measure

Development of the rating scale. The general purpose of the behavioral measure was to provide a vehicle for behaviorally assessing students' interpersonal styles when interacting with an adolescent. It was created by the researcher with the assistance of the staff of the Adolescent Diversion Project. Staff were asked to think of "critical incidents" involved in the current supervision of undergraduates with adolescents in the Project and then attempted to define behaviors which both capable and inept supervisees exhibited when working with the adolescents assigned to them. These initial behaviors broke into four general categories: style of obtaining information (level-headed, historical, judgmental, empathic, questionning style and patient); "purposefulness" (optimistic, planful, resourceful); perceived similarities with the youth (present and past) and nonverbal style (warm, anxious).

Table 5

Personality Research Form Component Structure

Component	Eigenvalue	Pct of Var	Cum Pct	Variable Name	Achievement Motivation
1	2.69491	19.2	19.2	achievement	.81550
2	2.28463	16.3	35.6	affiliation	.07880
3	1.99971	14.3	49.9	aggression	-.03201
4	1.51012	10.8	60.6	autonomy	.30745
5	.95570	7.0	67.6	dominance	.59524
6	.75161	5.4	73.0	endurance	.72606
7	.66358	4.7	77.7	exhibition	.32372
8	.61030	4.4	82.1	harm.-avoid.	-.30947
9	.54874	3.9	86.0	impulsivity	-.05066
10	.49491	3.5	89.5	nurturance	.19759
11	.42432	3.0	92.6	order	.15959
12	.39965	2.9	95.4	play	-.24195
13	.35230	2.5	97.9	social recog.	-.22563
14	.28953	2.1	100.0	understanding	.57881

Table 5 (Cont.)

Component	Impulsivity	Outgoingness	Aggression	Communality
1	-.12243	-.03633	.05154	.68401
2	.07600	.75798	-.17666	.61772
3	.11506	.03409	.79441	.64651
4	.40055	-.65176	.12028	.69422
5	.18704	.39579	.44759	.74628
6	-.07662	-.02307	-.08875	.54144
7	.37114	.56161	.34390	.67621
8	-.70653	-.02310	-.10385	.60628
9	.80087	.02245	.08614	.65188
10	.07995	.46064	-.57150	.58424
11	-.66153	.24231	.30641	.61568
12	.48598	.34563	.12483	.42976
13	-.16478	.57594	.30153	.50069
14	.22115	-.13252	-.30493	.49447

Table 6

Summary of PRF Components

Component	Items on Component	Descriptors of High Scorers
Achievement-Motivation	achievement endurance dominance understanding	striving, persistant controlling, analytical
Impulsivity	impulsivity not harmavoidance not order play	spontaneous, not fearful, not neat, playful
Outgoingness	affiliation not autonomy social recognition exhibition	friendly, not independent, approval-seeking, attention-seeking
Aggression	aggression not nurturance	argumentative, not sympathetic

Table 7

Intercorrelations of PRF Component Scores

	Achievement-Motivation	Impulsivity	Outgoingness	Aggression
Achievement-Motivation	1.00			
Impulsivity	.01	1.00		
Outgoingness	-.01	-.06	1.00	
Aggression	-.09	-.02	-.15*	1.00

*Significant at $p < .05$ ($N = 150$)

These characteristics were then placed on a five-point, bipolar, behaviorally anchored rating scale and assigned a category title.[1] In addition, overall evaluations sheets were created in a similar fashion for both participant raters (youth) and nonparticipant (observer) raters in order to tap their "hunches."

Development of situations. Following this, the staff created four situations which both reflected common situations occurring between the undergraduate and her/his assigned adolescent and could perhaps exhibit the categories of behaviors defined on the measure.

Training of role players. At this point in the development of the measure, two adolescents who had been referred to the Project and had either completed their commitment or were within several weeks of completing their commitments, were hired to be trained and work as both role players and participant raters. Each adolescent was assigned two of the four situations to present; the training as a role player consisted of cycles of the adolescent's interacting with staff members playing college students, presenting her/his two situations, to each, receiving feedback on her/his consistency of performance and interacting with another person. All of these interactions were tape recorded, in order

---

[1]Copies of all measures can be obtained by writing the author at the following address: Department of Psychology, Michigan State University, East Lansing, Michigan, 48824.

to serve as rater-training material. When role playing, the adolescents were given basic outlines of the problem to be presented and were supplied with an opening "line" to begin each situation. Details of what to say and how to react to people interviewing them were left up to the youth; this was decided upon because it was believed that it would be easier for the adolescents to be consistent across performances if they were able to draw from their own life experiences and feelings rather than attempting to memorize a set script. Adolescents met with the college role players for approximately seven hours of training.

Training raters. Once participant raters/role players had been trained in their role plays, attention turned to training both the participant and observer raters to rate the situations. Observer raters were two females who were formerly undergraduates involved with the Project and one female clerical Project staff member. Initially, both participant and observer raters met for a general introduction to the measure--an explanation and examples of the anchor points, a description of the rating process, etc. Next, all raters practiced rating segments from the audiotapes of an interaction, descussed their discrepant ratings, and rated another segment. After cycles of this, all raters then rated sequential segments of a taped interaction, discussed their discrepant ratings, and rated more segments. Thirdly, raters rated a complete five-minute interaction on audio-tape, still discussing discrepant ratings before rating another interaction. Finally,

the raters practiced rating "live" role plays similarly with staff role playing college students with the adolescents again.

Another important process was interspersed throughout this training. There was a continual revamping of the anchor points of the chracteristics. As raters discussed ratings, they helped to clarify and state the anchor points in more behavioral terms. Because of this process, it was decided to have the anchor points be frequency counts, with tally sheets for the observer raters only. (Since the participant raters were interacting face-to-face with the interviewees, it was thought that their using such tally sheets would be disruptive to the interviews.)

Once raters had attained interrater correlations of approximately .85, the creation of the measure moved into the piloting phase.

Piloting. The second portion of the training of raters took place during the official piloting of the measure. Two students from each training condition (Action, Relationship, Low Intensity--Large Group and Low Intensity--Small Group) were contacted to sign up for times to participate on a Tuesday evening. Time slots were 6:30, 6:40, 7:15, 7:25, 8:00, 8:00, 8:45, and 8:55. Thus, a person appeared for her/his appointed time and was greeted by the principal researcher. S/he was given a printed handout explaining the purpose of the measure and given some basic instructions on what to do. After the person had read the

sheet, s/he was escorted to one of the two interviewing rooms by a "manager." There, the manager introduced the student to the adolescent and gave more specific instructions as to the time flow of the interactions. Upon leaving the room, the manager set a timer for five minutes and turned on a tape recorder in the room. The adolescent handed the student an index card with background information about the first situation; when the student had read the information, s/he indicated so by looking up. The youth then began the interaction with her/his standard opening line for that situation.

If the student felt s/he had completed the interaction prior to the five-minute limit, s/he handed the information card back to the youth, and the manager (watching through the one-way mirror) came into the room. Otherwise, the manager reappeared at the end of the five minutes. Regardless of how the interaction ended, when the manager reappeared, the youth left the interviewing room and there was a one-minute break. During the break, the youth went to a nearby room to fill out a rating on that interaction; the raters remained in the viewing room to fill out ratings on that interaction, and the student "relaxed" in the interviewing room. After a sufficient time period for all to complete the forms (usually easily within one minute), the youth and the manager came back to the interviewing room; the student was given further (almost identical) instructions; the manager set the timer and the recorder was turned on; the student received another index card from the youth for the second situation; the student indicated

when s/he had read it; and the youth presented her/his standard opening line for the second situation. This interaction was terminated in either of the two ways mentioned above: either the student handed the card back to the youth if the five minutes had not transpired or the manager appeared when the timer went off. There again was the same break, with the youth leaving the room, the raters and the youth rating and the student remaining in the interviewing room. After this break, the manager appeared with the second adolescent, and the same procedure took place with that youth and her/his two situations.

After the second adolescent completed her/his second situation, the manager appeared and escorted both the youth and the college student from the interviewing room into the nearby room where the youth had been filling out her/his rating forms. As usual, the youth filled out a rating form on the last interaction and an overall rating form (having already filled out one on the first interaction during the prior break); however, the college student was then given five rating sheets to fill out: one for her/him to fill out on how s/he had behaved in each of the four situations and one to fill out on how s/he believed s/he "normally" interacted with adolescents. In this way, it allowed comparison of what the student believed to be her/his "normal" behavior with her/his actual performance as well as to compare her/his self-rating in each situation with what the raters saw. When the student completed all five forms, s/he was free to leave. The

complete process from the initial appearance to the final filling out of forms took approximately 45 minutes.

The timing of the assigned appointment times was crucial. A pair of students were assigned times close to each other in the following manner: one person showed up 10 minutes after the first one of the pair; the first student of the second pair then showed up 45 minutes after the time of the first student of the first pair, and so on. In this way, it was possible to standardize the order in which students talked with the adolescents. This was important, for the last interaction of one youth concerned some issues which could have carried over into other interactions if it were placed earlier in the order. Thus, as the first student of the pair was finishing up her/his interactions with the first youth, the second person was reading the handout explaining the measure. So, when the first youth had finished rating the second interaction s/he had just finished, s/he interacted with the second student of the pair while the second youth began her/his interactions with the first student. In this way, it was possible to minimize the amount of waiting for the adolescents while standardizing the order for the students.

Once piloted, the measure was run, with each student having an assigned time to appear, on a Tuesday or Wednesday evening or a Saturday morning for approximately three weeks in November. Interrater reliabilities were checked at random intervals during each session and feedback was given to raters at that time. In addition, feedback was given to the adolescents concerning the

consistency of their performance during each session in order to make certain that each subject was receiving as similar a presentation as was possible.

For each situation, then, there was a maximum of five ratings per category--three raters' scores, one adolescent score and a self score. In addition, there were three overall categories rated by any rater viewing all four situations, two overall rating categories filled out by each adolescent and a "self-usually" rating on the original 20 categories.

Development of factor scores. Once this myriad of scores was generated, it was necessary to attempt to collapse these data. First, endorsement frequencies were examined and low variance items were eliminated. Next, the 20 situation-specific categories were examined for generalizability of nonparticipant raters and intercorrelations of participant raters. Thirdly, a component structure underlying the 20 categories was sought which would cover all types of raters (observer, participant and self). After these analyses, the attention turned to the overall ratings by observer, participant and self raters. Intercorrelations within and across types of raters were examined; correlations of the "self-usually" ratings were compared with the component scores of self for the situations. Precise findings are reported in the following subsection.

Once the data had been obtained, endorsement frequencies were examined in order to eliminate any categories with little variance consistently across all types of raters. From this

process, four categories were dropped from further analysis: "illogical," "cold," "calm," and "anxious." Utilizing a FORTRAN program to calculate a coefficient of generalizability (Cronbach, Gleser, Nandy and Rajaratnam, 1972), interrater reliability of the observer raters was examined. The average coefficients of generalizability ranged from .51 (warm) to .96 (voiced similarities--then). (See Table 8.) Due to the high degree of convergence of these ratings, category scores were averaged across observer raters. Intercorrelations of participant raters' scores by category showed very little consistent and significant correlation across situation within rater and also very little consistent and significant correlation between raters. (See Appendix A.) Therefore, it was decided to do no further analyses utilizing the participant raters' scorings, due to the low probability of finding a meaningful factor structure.

After examining rater generalizability and correlations, focus turned to seeking an underlying component structure across situations within types of raters (observer and self). Observer rater scores' component structure seemed to vary from one situation to the next; however, self ratings exhibited an empirically as well as rationally sound structure. Again utilizing Kaiser's criteria, there were five components extracted: Voiced Social Comparison (similarities--now, similarities--then, dissimilarities--then, dissimilarities--now; loadings ranging from .77 to .67); Resourcefulness (planning, resourceful, optimistic--loadings ranging from .83 to .49); Judmentalness (judgmental, discounting,

Table 8

Average Generalizability Coefficients
for Observer Raters by
Characteristic

Characteristic	Average Generalizability Coeff.
optimistic	.76
pessimistic	.86
voiced similarities--now	.81
voiced dissimilarities--now	.80
voiced similarities--then	.96
voiced dissimilarities--then	.95
historical	.88
judgmental	.89
resourcful	.87
specific planning	.51
empathic	.68
discounting	.71
warm	.51
questioning style	.82
impatience	.55
talkative	.89

pessimistic--loadings ranging from .81 to .56); Warmth (warm, talkative, questioning style--loadings ranging from .67 to .62); Historicalness (historical--loading .77). (See Table 9.) In addition, the single item "Empathy" was retained as a singlet due to research interest. Because of the rational acceptability of this component structure, it was decided to apply this structure to the observer raters' scores as well.

Development of overall scores. Overall evaluations were examined by type of rater. Within each observer rater, all three overall categories were highly intercorrelated (quality of options, ease of interaction and likelihood to be effective change agent). Also, on the few occasions when two observer raters had overall evaluations on the same subjects, the three overall characteristic scores were highly intercorrelated. (See Appendix A.) Therefore, overall observer evaluations were collapsed both across overall characteristics and across raters to obtain one overall evaluation score. The participant overall evaluations (reaction to student, likelihood to help other youth) were significantly intercorrelated within rater (.52 and .97), but were not significantly correlated across rater (-.07 to .02). (See Table 10.) It was therefore decided to collapse the overall evaluations within participant rater, but not across participant raters. "Self--usually" component scores were significantly correlated with the component scores obtained by the self ratings of the four situations; the self component scores were therefore recomputed to include the

Table 9

Behavioral Measure Component Structure
(Self Rating)

Component	Eigenvalue	Pct of Var	Cum Pct	Variable Name
1	2.91214	18.2	18.2	optimistic
2	2.28154	14.3	32.5	pessimistic
3	1.64565	10.3	42.7	similarities-now
4	1.29735	8.1	50.9	dissimilar-now
5	1.12774	7.0	57.9	similarities-then
6	.98221	6.1	64.0	dissimilar-then
7	.86263	5.4	69.4	historical
8	.81006	5.1	74.5	judgmental
9	.68371	4.3	78.8	resourceful
10	.63165	3.9	82.7	planning
11	.55192	3.4	86.2	empathic
12	.53398	3.3	89.5	discounting
13	.51291	3.2	92.7	warm
14	.48153	3.0	95.7	questioning style
15	.36031	2.3	98.0	impatience
16	.32467	2.0	100.0	talkative

Table 9 (Cont.)

Component	Voiced Social Comparison	Resource-fulness	Judgmen-talness	Warmth	Histori-calness	Commun-nality
1	.25880	.49176	.00339	.21969	.45953	.56825
2	.04254	-.10285	.73660	.10667	-.02123	.56680
3	.76947	.10442	-.01139	.01594	.10628	.61467
4	.66548	-.00810	.29495	-.10663	-.02823	.54209
5	.73610	.02799	-.20512	.24845	.20006	.68645
6	.70344	.09200	.14787	.10375	-.33684	.64939
7	-.04759	.05098	-.00970	.10470	.76603	.60272
8	.03794	.01861	.81566	-.04205	.06400	.67296
9	.10072	.71166	-.00103	.17000	.05677	.54873
10	-.08305	.82856	.05717	.17212	-.02129	.72675
11	.19746	.40421	-.24729	-.37731	.34417	.52434
12	.08330	.33846	.55731	-.14336	-.18396	.48649
13	.12308	.24107	-.14890	.67408	.17122	.57913
14	.10638	.28767	.00298	.61886	-.22177	.52624
15	.37705	-.24688	.38346	-.20172	.33794	.50505
16	-.01751	.01234	.03244	.66477	.14463	.46436

Table 10

Intercorrelations of Participant Overall Evaluation Component Scores

	Rater 1 Char. 1	Rater 1 Char. 2	Rater 2 Char. 1	Rater 2 Char. 2
Rater 1 (like/dislike student)				
Rater 1 (helpful to other youth)	.52			
Rater 2 (like/dislike student)	.02	-.07		
Rater 2 (helpful to other youth)	-.04	.00	.97	

"self-usually" scores prior to averaging with each component.

Thus, from the behavioral measure there emerged 15 average component scores: six for observer raters (Voiced Social Comparison, Resourcefulness, Judgmentalness, Warmth, Historicalness, Empathy); six for self ratings (Voiced Social Comparison, Resourcefulness, Judgmentalness, Warmth, Historicalness, Empathy); two Overall Evaluations for participant raters; and one Overall Evaluation for observer raters. Table 11 offers a summary of the final component scores; Table 12 offers the intercorrelations of the final component scores.

Peer, Supervisor and Self In-Class Evaluation Rankings

Measure development. These measures were created in order to assess what students, their peers and their supervisors perceived individuals to be like during supervision sessions. These measures were examined in order to determine if they had any predictive value when dealing with outcome--in other words, does what a student's peers and supervisors (as well as the student her/himself) think of her/him on such traits as "comfortable with the approach," "confident," etc., correlate with whether her/his assigned youth "succeeds" in the program. Also, the relationships of these measures to other measures were scrutinized. (The supervisor evaluations differ from the peer evaluations in title only: both the peers, the supervisors and the students themselves rate and rank the students on identical dimensions.)

Table 11

Summary of Behavioral Measure Components

Component Name	Items on Component	Description
Observer Raters		
Voiced Social Comparison	similarities--now dissimilarities--then similarities--then dissimilarities--then	Student compared self and adolescent as rated by observer raters
Resourcefulness	resourceful specific planning optimistic	Students thought of different ideas, steps to take and encouraged youth, as rated by observer raters
Judgmentalness	judgmental discounting pessimistic	Student voiced opinions, discounted feelings and discouraged adolescent, as rated by observer raters
Warmth	warm talkative questioning style	Student was nonverbally warm, talked more and asked more openended questions, as rated by observer raters
Historicalness	historical	Student asked about adolescent's past, as rated by observer raters
Empathy	empathic	Student labeled adolescent's feelings, as rated by observer raters
Self		
Voiced Social Comparison	similarities--now dissimilarities--now similarities--then dissimilarities--then	Student compared self and adolescent as rated by self

Table 11 (Cont.)

Component Name	Items on Component	Description
Resourcefulness	resourceful specific planning optimistic	Student thought of different ideas, steps to take and encouraged youth, as rated by self
Judgmentalness	judgmental discounting pessimistic	Student voiced opinions, discounted feelings and discouraged youth, as rated by self
Warmth	warm talkative questioning style	Student was nonverbally warm, talked more and asked more openended questions, as rated by self
Historicalness	historical	Student asked about youth's past, as rated by self
Empathy	empathic	Student labeled youth's feelings, as rated by self
Overall Evaluations		
Overall evaluation, observer rater	quality of options ease of interaction likelihood to be effective change agent	Overall evaluation by observer raters
Overall evaluation, participant rater 1	reaction to student likely to help other youth	Overall evaluation by participant rater 1
Overall evaluation,	reaction to student likely to help other youth	Overall evaluation by participant rater 2

Table 12

Intercorrelations of Behavioral Measure Component Scores

	R1	R2	R3	R4	R5	R6	S1	S2	S3	S4	S5	S6	OVR	PR1	PR2
Observer Raters															
Voiced Social Comparison (R1)	.48														
Resourcefulness (R2)	.65	.46													
Warmth (R4)	.14	.15	.04												
Historicalness (R5)	-.22	.05	-.04	-.35											
Empathy (R6)	-.12	.02	.01	-.40	.23										
Self															
Voiced Social Comparison (S1)	.18	-.08	.19	.18	-.30	-.02									
Resourcefulness (S2)	-.13	.20	.01	.02	.12	-.11	.41								
Judgmentalness (S3)	-.10	-.05	.19	.18	-.30	-.11	.30	.22							
Warmth (S4)	-.14	.03	.07	.04	.13	-.17	.18	.54	-.07						
Historicalness (S5)	-.25	-.02	-.22	.11	.22	.00	.13	.35	.05	.24					
Empathy (S6)	-.16	-.08	-.18	-.13	.11	.30	.10		.32	-.10	.17	.30			
Overall Evaluations															
Overall Evaluation, observer rater (OVR)	-.38	.27	-.25	.50	-.29	-.11	.23	.42	.18	.11	.27	.18			

Table 12 (Cont.)

	R1	R2	R3	R4	R5	R6	S1	S2	S3	S4	S5	S6	OVR	PR1	PR2
Overall Evaluation, participant rater 1 (PR1)	-.66	-.34	-.57	.17	.05	-.13	.02	.19	-.08	.28	.40	.26	.35		
Overall Evaluation, participant rater 2 (PR2)	-.47	-.29	-.31	.11	.17	-.05	.04	.10	.01	.07	.05	.05	.31	.33	

Peer and supervisor evaluations were created by the staff of the Adolescent Diversion Project. Characteristics consisted of those that undergraduates had exhibited during supervision sessions, as perceived by staff members who had worked in such supervisory sessions. Characteristics were rated on a scale from one to five, with anchor points of the type "is always like this" to "is never like this." Again, characteristics could be grouped into larger categories: style of applying class skills (e.g., confidence, comfortableness, effectiveness, planfulness, perseverance); ability to adapt to assigned youth and significant others (e.g., getting along with youth, family and school officials); style of interaction in class (understanding the model, talkativeness, following through with suggestions, giving input into others' cases, showing interest, helpfulness); and general impressions (intelligence, motivation, success in career, well-liked).

Each student rated both her/himself and all other undergraduates in her/his supervision group. (Most groups consisted of either six or seven supervisees and two staff supervisors.) In addition to rating all supervisees, each student ranked the same students on the same characteristics (again including her/himself). Supervisors, likewise, rated and ranked all of the undergraduate supervisees in their groups. Thus, for each person at each time period, there were from five to 14 peer ratings and rankings (depending on section size), two supervisor ratings and rankings, and one self rating and ranking for each of the characteristics. These measures were collected at three points

during the three-term course--at the end of the training sequence (eight weeks); at the midpoint from the end of training to the end of the three terms (18 weeks) and at the end of the course sequence (30 weeks).

Development of factor scores. Due to the probability of social desirability influencing students' ratings of themselves and their peers, it was decided to focus primarily on rankings. In scrutinizing the peer, supervisor and self rankings, two major areas were focused upon. First, generalizability of peer rankings were examined across raters and intercorrelations of supervisor rankings were examined. Following this, attempts were made to ferret out an underlying component structure within and across the three types of raters (peer, supervisor and self). These results are presented in this subsection.

First, coefficients of generalizability were computed for peer rankings completed at the first time period (end of training) within sections. Where computable, average coefficients were generally high, ranging from .53 (empathic) to .91 (talks in supervision sessions). (See Table 13.) Because of this, it was decided to collapse peer rankings across raters and assign an average characteristic score to each of the 21 characteristics. Intercorrelations of supervisor rankings were computed; again, due to high correlations between the two supervisors, supervisor rankings were averaged by characteristics. (See Table 13.)

Table 13

Generalizability of Peer Rankings and Intercorrelations of Supervisor Rankings In-Class Evaluation Rankings

Characteristic	Average Peer Rater Generalizability Coeff.	Supervisor Rankings Intercorrelations
confident	.91	.75
agrees with approach	.63	.84
effective	.76	.70
understands model	.75	.42
talkative	.91	.63
gets along with youth	.71	.70
gets along with family	.66	.45
gets along with school officials	.61	.33
planful	.58	.55
perseverant	.62	.73
resourceful	.83	.48
motivated	.77	.79
successful in career	.68	.50
well-liked by supervisors	.64	.57
well-liked by students	.71	.36
empathic	.53	.17
warm	.59	.81
follows through	.54	.36
gives input	.80	.73
shows interest	.77	.55
helpful	.82	.53

These average characteristic scores were then examined by type of rater (peer, supervisor and self) using a principal components analysis with varimax rotation. All three types of raters exhibited the same pattern. Using Kaiser's criteria, clearly component accounted for the majority of the variance in all the scores: peer rankings, 86%, supervisor rankings, 77%, self rankings, 77%. (See Tables 14 to 16.) Thus, within each time period, there were computed three average rank scores by adding all of the ranks and dividing by 21 for each type of rater. Therefore, at this point, there were nine average rank scores--three for time period one (end of training); three for time period two (18 weeks from beginning the Project); and three for time period three (end of the Project). These nine rank scores were then converted to T-scores (Guilford, 1942) due to differing Ns per section (from 6 to 15). However, upon examining the correlation matrix of these nine measures as T-scores,it was discovered that these nine scores were also highly intercorrelated (as shown in Table 17.) Therefore, the nine scores were again averaged to obtain one measure of overall in-class evaluation ranking.

Process Measures

Measure development. Process measures were examined in order to ascertain whether what occurred between the student and her/his assigned youth--how they got along, how often they got together, etc., as reported by different sources--in any way predicted to outcome or any other measures. For example, it was

Table 14

In-Class Evaluation Rankings by Peers

Component	Eigenvalue	Pct of Var	Cum Pct	Variable Name	Evaluation Component	(Component 2)	Communality
1	18.08429	86.1	86.1	confident	.86493	.41852	.92326
2	1.19077	5.7	91.8	agrees w/approach	.40786	.83108	.85705
3	.36146	1.7	93.5	effective	.58378	.74652	.89809
4	.26139	1.2	94.8	understands model	.71320	.65595	.93893
5	.21164	1.0	95.8	talkative	.90580	.35555	.94689
6	.18592	.9	96.6	gets along w/youth	.62079	.69562	.86927
7	.15615	.7	97.4	gets along w/family	.60119	.69218	.84054
8	.10946	.5	97.9	gets along w/school officials	.47814	.81157	.88726
9	.07899	.4	98.3	planful	.41497	.82146	.84699
10	.06454	.3	98.6	perseverant	.60199	.72640	.89004
11	.05397	.3	98.9	resourceful	.88338	.43019	.96542
12	.04498	.2	99.1	motivated	.87489	.43797	.95724
13	.03996	.2	99.3	successful in career	.73896	.62869	.94132
14	.03855	.2	99.4	well-liked by supervisors	.72151	.65332	.94741
15	.02844	.1	99.6	well-liked by students	.69871	.67438	.94299
16	.02375	.1	99.7	empathic	.59646	.75457	.92514
17	.02123	.1	99.8	warm	.55090	.78978	.92724
18	.01708	.1	99.9	follows through	.23397	.91627	.89429
19	.01043	.0	99.9	gives input	89482	.38717	.95061
20	.00906	.0	100.0	shows interest	.81617	.54358	.96162
21	.00794	.0	100.0	helpful	.82967	.52452	.96347

Table 15

In-Class Evaluation Rankings by Supervisors

Component	Eigenvalue	Pct of Var	Cum Pct	Variable Name	Evaluation Component	(Component 2)	Communality
1	16.19160	77.1	77.1	confident	.61593	.71032	.88392
2	.94342	4.5	81.6	agrees w/approach	.81820	.27312	.74405
3	.83578	4.0	85.6	effective	.80922	.49858	.90342
4	.62167	3.0	88.5	understands model	.81476	.05814	.66725
5	.49185	2.3	90.9	talkative	.50487	.74468	.80944
6	.38296	1.8	92.7	gets along w/youth	.62743	.64478	.80942
7	.27965	1.3	94.0	gets along w/family	.59021	.63614	.75302
8	.25610	1.2	95.3	gets along w/school officials	-.01224	.73622	.54216
9	.21308	1.0	96.3	planful	.70403	.52466	.77093
10	.17373	.8	97.1	perseverant	.74414	.53997	.84532
11	.12748	.6	97.7	resourceful	.63034	.68514	.86675
12	.09197	.4	98.1	motivated	.60786	.72740	.89860
13	.08139	.4	98.5	successful in career	.71127	.61432	.88329
14	.07497	.4	98.9	well-liked by supervisors	.76937	.49911	.84103
15	.06646	.3	99.2	well-liked by students	.73100	.50988	.79433
16	.04639	.2	99.4	empathic	.69011	.62148	.83504
17	.04117	.2	99.6	warm	.66993	.62148	.83504
18	.02728	.1	99.7	follows through	.77612	.48588	.83844
19	.02214	.1	99.9	gives input	.57414	.76409	.91347
20	.01843	.1	99.9	shows interest	.59015	.73728	.89186
21	.01247	.1	100.0	helpful	.60220	.66560	.80567

Table 16

Component	Eigenvalue	Pct of Var	Cum Pct	Variable Name	Evaluation Component	(Component 2)	Communality
1	16.18247	77.1	77.1	confident	.72279	.58851	.86878
2	1.31416	6.3	83.3	agrees w/approach	.29091	.83282	.77822
3	.87219	4.2	87.5	effective	.66878	.62124	.83321
4	.52391	2.5	90.0	understands model	.66059	.60633	.80405
5	.37301	1.8	91.7	talkative	.83282	.40571	.85818
6	.32841	1.6	93.3	gets along w/youth	.54404	.74788	.85531
7	.30606	1.5	94.8	gets along w/family	.52640	.73293	.81429
8	.23057	1.1	95.9	gets along w/school officials	.41800	.80606	.82446
9	.17492	.8	96.7	planful	.19384	.86249	.78146
10	.15588	.7	97.4	perseverant	.50153	.79192	.87866
11	.10507	.5	97.9	resourceful	.74712	.41607	.73130
12	.08805	.4	98.4	motivated	.76004	.49156	.81928
13	.07710	.4	98.7	successful in career	.62075	.70174	.87777
14	.05947	.3	99.0	well-liked by supervisors	.83284	.45054	.89662
15	.04920	.2	99.2	well-liked by students	.86108	.40798	.90791
16	.04105	.2	99.4	empathic	.77713	.48328	.83749
17	.03357	.2	99.6	warm	.74060	.51626	.81502
18	.02693	.1	99.7	follows through	.74505	.50147	.80657
19	.02533	.1	99.8	gives input	.84250	.17936	.74197
20	.01958	.1	99.9	shows interest	.71192	.59314	.85864
21	.01309	.1	100.0	helpful	.90266	.30442	.90747

Table 17

Intercorrelations of Peer, Supervisor and Self In-Class Evaluations Rankings Component Scores

	P1	Su1	Se1	P2	Su1	Se2	P3	Su3	Se3
Peer Ranking, Time 1 (P1)									
Supervisor Ranking, Time 1 (Su2)	.83								
Self Ranking, Time 1 (Se1)	.60	.58							
Peer Ranking, Time 2 (P2)	.74	.65	.58						
Supervisor Ranking, Time 2 (Su2)	.57	.66	.41	.75					
Self Ranking, Time 2 (Se2)	.50	.36	.56	.51	.35				
Peer Ranking, Time 3 (P3)	.59	.58	.50	.82	.71	.33			
Supervisor Ranking, Time 3 (Su3)	.55	.58	.37	.75	.90	.29	.79		
Self Ranking, Time 3 (Se3)	.47	.51	.45	.40	.44	.48	.46	.48	

possible to look at whether the youth's reporting that s/he liked her/his assigned student was related to the youth's staying out of trouble while s/he was involved with the assigned college student. This information was obtained from the "intervention" questions of the process interviews conducted during the intervention period. Table 18 shows the process interview schedule.

Development of scale scores. Interviews were conducted with the adolescent referred to the Project, her/his parent, one peer nominated by the adolescent and the undergraduate assigned to the youth, as mentioned earlier, by interviewers blind to any student measures. Endorsement frequencies were examined, and any questions with little variance were dropped. Scales of the remaining items were constructed using a rational-empirical method (Jackson, 1970). Fourteen scales emerged from this procedure, with Cronbach's alphas ranging from .51 (Recreational Activities) to .95 (Contracting Activities). (See Table 19.) Due to high correlations of scale across scoures of information, it was decided to collapse the scale scores across source. Thus, there were 14 scale scores at each of three time periods, or 42 scale scores available for analyses. Intercorrelations of the final scale scores within time period are presented in Tables 20 to 22.

Outcome Measures

Measure development. Outcome criteria were what had happened with the youth during the intervention period. Outcome was obtained from three sources initially: police records

Table 18

Process Interview Schedule

	Assignment to Project	After six weeks	After twelve seeks	Termination
Assigned youth	X[a]	X	X	X
Parent	X[a]	X	X	X
Peer nominated by youth	X[a]	X	X	X
Assigned undergraduate student		X	X	X

[a]Interview does not include "Intervention" questions at this point; will not be used in this research.

Table 19

Summary and Reliability of Intervention Scale Scores

---

1. Volunteer/Target Involvement: frequency and amount of contact. (α = .53)

2. Lack of Complaints/Positive Involvement: the extent to which the youth and the assigned volunteer get along and the lack of problems involved in the intervention process. (α = .86)

3. Parent Involvement: the extent to which parent(s) are included in the intervention and the extent of a relationship built up between the parent(s) and the volunteer. (α = .85)

4. Peer Involvement: extent to which friends of the youth are involved in the intervention. (α = .81)

5. Recreational Activity: amount of recreation involved in the time spent with youth by volunteer. (α = .51)

6. Family: Focus on Changing Parents: extent to which the intervention focused upon changing the parents' behavior in the family. (α = .86)

7. Family: Focus on Changing Youth: extent to which the intervention focused upon changing the youth within the family context. (α =.71)

8. School: Focus on Changing School: extent of the intervention focusing on bringing improvement to the school environment by focusing on school staff. (α = .70)

9. School: Focus on Changing Youth: extent of the intervention focusing on school behavior of the youth. (α = .86)

10. Job-seeking: extent to which the intervention focused upon getting the youth employment. (α = .86)

11. Legal System Involvement: extent to which the volunteer became involved in the juvenile justice system for the youth. (α = .77)

12. Contracting Activities: extent to which the voluntter utilized behavioral contracting as an option in the intervention. (α = .95)

Table 19 (Cont.)

---

13. Relationship Activities: extent to which the volunteer utilized relationship building as an intervention strategy. (α = .84)

14. Advocacy Activities: extent to which the volunteer intervened on behalf of the youth to gain needed resources. (α = .82)

Table 20

Intercorrelations of Intervention Scale Scores, Time 1

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14
Vol./Target Involvement (I1)														
Lack of Complaints/+ Involvement (I2)	.12													
Parental Involvement (I3)	.42	.00												
Peer Involvement (I4)	-.03	.00	21											
Recreational Activities (I5)	.06	.18	-.09	-.08										
Family: Focus on Parents (I6)	-.05	-.23	.39	.22	-.21									
Family: Focus on Youth (I7)	-.10	-.13	.18	.29	-.25	.73								
School: Focus on School (I8)	.05	-.09	.15	-.01	-.25	.21	.19							
School: Focus on Youth (I9)	-.06	-.13	.43	.08	-.10	.28	.40	.58						
Job-seeking (I10)	.07	.05	.20	-.06	-.07	.32	.17	.20	.07					
Legal System Involvement (I11)	-.11	-.04	.18	.20	-.29	.33	.21	.25	.14	.28				
Contracting Activities (I12)	.00	.14	.21	.09	-.02	.37	.40	.41	.36	-.03	-.06			
Relationship Activities (I13)	.01	-.02	.17	-.07	-.15	.34	.46	.36	.45	.39	.21	.20		
Advocacy Activities (I14)	.01	-.09	.26	.03	-.06	.35	.30	.63	.53	.56	.31	.39	.66	

Table 21

Intercorrelations of Intervention Scale Scores, Time 2

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14
Vol./Target Involvement (I1)														
Lack of Complaints/+ Involvement (I2)	.55													
Parental Involvement (I3)	.35	.29												
Peer Involvement (I4)	.21	.11	.29											
Recreational Activities (I5)	.51	.25	.07	.25										
Family: Focus on Parents (I6)	.12	.13	.57	.28	-.05									
Family: Focus on Youth (I7)	.00	.03	.39	.18	.02	.73								
School: Focus on School (I8)	.04	.04	.37	.10	-.04	.38	.27							
School: Focus on Youth (I9)	.11	.15	.53	.11	.16	.45	.61	.53						
Job-seeking (I10)	.27	.23	.32	.04	-.08	.37	.24	.00	.19					
Legal System Involvement (I11)	.07	.02	.31	.11	-.02	.35	.33	.51	.39	.14				
Contracting Activities (I12)	.12	.22	.39	.06	.05	.45	.32	.62	.39	.18	.32			
Relationship Activities (I13)	.27	.21	.30	.19	.02	.53	.61	.31	.55	.47	.21	.29		
Advocacy Activities (I14)	.32	.28	.36	.24	.14	.56	.42	.41	.52	.62	.25	.39	.69	

Table 22

Intercorrelations of Intervention Scale Scores, Time 3

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14
Vol./Target Involvement (I1)														
Lack of Complaints/+ Involvement (I2)	.59													
Parental Involvement (I3)	.50	.31												
Peer Involvement (I4)	.27	-.01	.40											
Recreational Activiites (I5)	.45	.36	.12	.16										
Family: Focus on Parents (I6)	.09	.06	.51	.42	-.11									
Family: Focus on Youth (I7)	.05	.11	.37	.31	.09	.81								
School: Focus on School (I8)	.07	.27	.37	.08	-.01	.30	.30							
School: Focus on Youth (I9)	.05	.34	.45	.22	.15	.41	.57	.61						
Job-seeking (I10)	.19	.07	.32	.18	-.01	.41	.57	.61						
Legal System Involvement (I11)	.16	.04	.23	.24	.21	.19	.30	.43	.43	.09				
Contracting Activities (I12)	.15	.10	.27	.14	.19	.30	.34	.38	.36	.07	.56			
Relationship Activities (Ile)	.17	.26	.24	.27	.12	.55	.62	.17	.45	.25	.16	.03		
Advocacy Activities (I14)	.10	.19	.28	.35	.04	.55	.52	.51	.50	.56	.48	.32	.59	

(number and severity of contacts), court records (number and severity of court contacts) and school records (grade point average, number of days present, number of credits earned). All data were collected by coders blind to any of the student measures.

Development of outcome score. Once the outcome data had been collected, it was scrutinized for meaningful patterns. At this point, it was decided to ignore school data, because there was no clear breakoff between attenders and nonattenders. Ideally, police contacts would be used as the sole measure of recidivism (failure). However, due to the incompleteness of available police data, it was decided to classify as a failure anyone having any further contact with either a police department or the Probate Court. In this way, the most conservative criterion was adopted, for even the most minor offenses were considered a failure on the part of the Project.

CHAPTER III

RESULTS

From the various scale and component score developments reported above, there were 62 independent variables remaining at this stage: 15 component scores of the behavioral measure, four PRF components, one overall ranking based on in-class evaluations and 14 Intervention scale scores at each of three time periods. In this section, results of three major areas of interest in this research will be examined. The effect of different training conditions on the behavioral measure will be determined using analyses of variance and planned comparison Scheffé tests. The interrelationship of the potential predictors of a successful volunteer will also be scrutinized by examining the intercorrelations of the independent measures. Finally, the results of the major design will be reported--looking at the multivariate predictive power of the independent measures and the potential building of a selection battery to forecast a "successful" volunteer.

Effects of Training Conditions on Behavioral Measure

Due to the lateness of the administration of the behavioral measure (after approximately 3/4 of the training was completed), it was conceivable that the training conditions could have interacted

with the behavioral measure factor scores. Therefore, a step-down multivariate analysis of variance was computed, examining the effects of the 15 component scores of the behavioral measure as classified by type of training received. Since a significant multivariate F-ratio was obtained ($F = 2.16$, $p < .001$), univariate oneway analyses of variance were conducted on each component score by type of training. Six component scores yielded significant results at the $p < .05$ level. Tables 23 to 28 display these results; the following subsections describe these results in detail.

Observer raters. Four component scores as scored by the observer raters were significant. Using a planned comparison Scheffé test, it was discovered that the Action training condition was rated significantly more highly by observer raters on Resourcefulness than the other three conditions ($F = 4.08$, $p < .05$). Students trained in this condition thought of more plans and steps for their ideas for the adolescents. This was anticipated, since the Action training condition focused primarily on assessing the youth's needs, brainstorming options and generating specific plans to follow in carrying out the options. Secondly, students in the Low Intensity--Large Group training condition were seen as significantly more judgmental by the observer raters than the other three training conditions ($F = 10.65$, $p < .05$). Due to the infrequency and lower intensity of supervision received by these students, it was suspected that perhaps these students would not be "schooled" as well as the other students in thinking objectively

Table 23

Analysis of Variance

Resourcefulness (Observer Rater)
by Training Condition

Source	Df	Ms	F	Sign.	ω^2
Between conditions	3	.58	3.91	.013	.13
Within conditions	53	.15			

Table 24

Analysis of Variance

Judgmentalness (Observer Rater)
by Training Condition

Source	Df	Ms	F	Sign.	ω^2
Between conditions	3	.54	5.28	.002	.18
Within conditions	53	.10			

Table 25

Analysis of Variance

Empathy (Observer Rater)
by Training Condition

Source	Df	Ms	F	Sign.	ω^2
Between conditions	3	1.57	4.60	.006	.16
Within conditions	53	.34			

Table 26

Analysis of Variance

Voiced Social Comparison (Self)
by Training Condition

Source	Df	Ms	F	Sign.	ω^2
Between conditions	3	.46	3.63	.0-9	.12
Within conditions	53	.13			

Table 27

Analysis of Variance

Judgmentalness (Self) by
Training Condition

Source	Df	Ms	F	Sign.	ω^2
Between conditions	3	.71	3.51	.021	.12
Within conditions	53	.20			

Table 28

Analysis of Variance

Overall Evaluation (Observer Rater)
by Training Condition

Source	Df	Ms	F	Sign.	ω^2
Between conditions	3	1.77	2.98	.04	.10
Within conditions	53	.60			

about their youth. Thirdly, students receiving the Relationship training were rated as more empathic (i.e, labeling more of the youth's feelings) than any of the other conditions by the observer raters ($F = 9.46$, $p < .05$). This was clearly the way in which they were trained to interact with their youth, and apparently they did so significantly more often than students from the other three conditions. Finally, on the Overall Evaluation by the observer raters, those students receiving the Action training were rated more highly by the observer raters ($F = 6.76$, $p < .05$). This implies that the observer raters felt that the students in the Action training condition were more likely to speak easily with the youth, to create more helpful options and to eventually be a more effective volunteer. This was anticipated for two reasons. First, although the observer raters had been trained on the frequency count items, the three overall evaluation items were left up to the observer raters' hunches. Although all of the observer raters were blind to condition, two of the three were at one time trained in the Action condition. It was therefore possible that if there was a difference among conditions, the observer raters would favor students on the Overall Evaluation than the other three conditions. Students in the Relationship skills training condition were rated by observer raters more highly on Empathy, and rated themselves as lower on both the Voiced Social Comparison and the Judgmentalness scores than the other three training conditions. The students in the Low Intensity--Large

Group were rated by observer raters as higher on the Judgmentalness score; the students in the Low Intensity--Small Group, however, appeared to have no distinguishing features on the behavioral measure.

Intercorrelations of Independent Measures

At this point in the analyses, the number of subjects in the study changed slightly. Two students were assigned adolescents, but due to unavailability of the assigned youth for process data (one youth ran away and one youth decided to drop from the Project after assignment), there is complete data for the following analyses on only 53 subjects.

Relationships between behavioral measure and PRF. Of the 15 component scores of the behavioral measure, only one was significantly correlated ($p < .05$) with any of the four PRF components. Namely, the self-rating of Voiced Social Comparison was correlated with Impulsivity on the PRF (-.29). Those who rated themselves as having mentioned similarities and dissimilarities with the youth more often were less impulsive and playful, according to the PRF factor score. So, of 60 possible correlations, only one was significant.

Relationships between behavioral measure and overall in-class evaluation rank. Similarly, one one behavioral measure score was correlated significantly ($p < .05$) with the overall in-class evaluation rank. Here, the self-rating of Historicalness was

correlated with the in-class evaluation rank. Students who believed that they asked questions about the youth's past in the behavioral measure were rated more highly in the in-class evaluation ranking. Thus, out of the possible 15 correlations, only one was significant.

Relationships between the behavioral measure and the Intervention Scale. Tables 29 to 31 present the intercorrelations of the behavioral measure component scores with the Intervention scale scores at each of the three time periods.

At time 1, 24 of the 210 correlations were significant (Table 29). Students rated by the observer raters as more resourceful (generating more plans and steps to plans and more optimistic) also focused more on both the school and the youth in the school context, did more contracting, relationship and advocacy activities. Students rated as warmer by the observer raters complained more and were positively involved with their youth less while being more involved with the parents. Those students rating themselves as voicing more similarities and dissimilarities with the youth also complained more and were less positively involved with their youth and engaged in more advocacy activities. Students rating themselves as more resourceful also exhibited a greater focus on changing both the school and the youth in school and greater relationship and advocacy activities. Those rating themselves as more judgmental in the behavioral measure engaged in more recreational activities with their youth. People who rated themselves as more historical

Table 29

Intercorrelations of Behavioral Measure Component Scores with Intervention Scale Scores, Time 1

	R1	R2	R3	R4	R5	R6	S1	S2	S3	S4	S5	S6	OVR	PR1	PR2
I1	-.21	-.07	.04	.03	-.04	.14	-.10	.05	.15	.05	-.15	-.05	.07	.14	.00
I2	-.08	.13	.04	-.28	.10	.12	-.33	.03	-.06	.14	-.04	.04	-.02	-.24	-.16
I3	-.10	.18	-.08	.28	-.17	-.23	.02	.10	.09	.07	.14	-.15	.34	.00	-.06
I4	.08	-.03	.04	-.01	-.14	-.16	.04	-.23	-.11	.07	.12	-.19	-.14	.02	-.04
I5	-.08	-.13	.14	.07	-.07	.18	.14	.16	.27	.05	-.02	-.06	-.09	-.14	.04
I6	-.02	.07	-.07	.19	-.04	-.20	.13	.05	-.01	-.01	.24	.11	.26	.06	.10
I7	.07	.26	-.01	.19	-.03	-.10	.04	.02	.03	-.09	.26	.17	.34	.05	.09
I8	-.07	.32	-.01	.03	-.07	-.11	.08	.31	.05	.23	.28	-.08	.35	.10	-.18
I9	.08	.45	.02	.25	-.07	-.18	.09	.34	.08	.24	.30	-.05	.33	.03	-.15
I10	-.10	.04	-.01	-.01	-.02	-.01	.08	.04	-.14	.09	.25	.08	.26	-.05	.18
I11	-.03	.01	.01	.03	-.12	-.01	.21	.02	.00	.16	.37	.25	.17	.11	-.06
I12	-.02	.28	.01	.07	.08	-.14	-.04	.19	.14	.01	.23	.11	.22	-.20	.03
I13	.01	.36	-.03	.19	.07	.03	.10	.36	-.08	.14	.15	.21	.42	-.03	.07
I14	-.02	.29	-.08	.17	-.08	-.06	.31	.44	-.04	.25	.44	.18	.47	.02	.05

Key:

Intervention Scale

I1:Vol./Target Inv.
I2:Lack of Complaints
I3:Parental Inv.
I4:Peer Inv.
I5:Rec. Activities
I6:Family: Parents
I7:Family: Youth
I8:School: School
I9:School: Youth
I10:Job-seeking
I11:Leg. Sys. Inv.
I12:Contracting Agt.
I13:Rel. Activities
I14:Advocacy Act.

Behavioral Measure

Observer Raters

R1:V. Soc. Comparison
R2:Resourceful.
R3:Judgmental
R4:Warmth
R5:Historical
R6:Empathy

Self

S1:V. Soc. Comparison
S2:Resourceful.
S3:Judgmental.
S4:Warmth
S5:Historical.
S6:Empathy

Overall evaluation

OVR: observer raters
PR1: participant rater 1
PR2: participant rater 2

Table 30

Intercorrelations of Behavioral Measure Component Scores with Intervention Scale Scores, Time 2

	R1	R2	R3	R4	R5	R6	S1	S2	S3	S4	S4	S6	OVR	PR1	PR2
I1	-.12	-.13	.09	-.14	.06	.06	-.10	-.07	-.02	.07	-.18	-.01	.00	-.18	.13
I2	-.11	.03	.07	-.19	.03	.07	-.14	-.05	-.11	.06	-.13	-.04	-.02	-.30	-.10
I3	-.14	.13	-.06	.15	-.08	-.15	.05	.24	.10	.02	.09	.00	.31	-.03	.02
I4	-.11	-.08	-.15	.25	-.20	-.11	.04	-.18	-.17	.07	.14	-.13	.06	.10	.09
I5	.17	-.16	.25	-.06	-.02	.07	.21	.04	.07	.05	-.20	-.08	-.23	-.23	.04
I6	-.05	.06	-.05	.27	.00	-.10	.25	.21	.01	.03	.36	.13	.36	.03	.11
I7	.09	.27	.02	.36	-.08	-.05	.21	.31	-.01	.02	.29	.27	.42	-.02	.10
I8	.09	.10	.00	.06	-.14	-.18	.15	.18	.03	.09	.10	-.01	.16	-.16	.12
I9	.13	.36	.16	.26	-.13	-.10	.28	.37	.00	.18	.21	.06	.31	-.08	.00
I10	-.07	.09	.03	.00	-.15	-.09	.17	.15	.05	-.02	.02	-.01	.28	-.15	.09
I11	.04	.17	.12	-.03	-.06	-.16	-.03	.15	-.13	.02	.08	-.05	.18	-.02	-.04
I12	-.06	.26	-.01	.07	.02	-.20	-.07	.20	.11	.00	.13	.05	.25	-.18	.03
I13	.01	.28	.00	.25	.01	.10	.04	.27	-.16	.11	.12	.23	.34	-.03	.15
I14	-.08	.12	.04	.09	-.14	.09	.35	.33	.12	.14	.19	.22	.37	-.09	.12

Key:

Intervention Scale

I1:Vol./Target Inv.
I2:Lack of Complaints
I3:Parental Inv.
I4:Peer Inv.
I5:Rec. Activities
I6:Family: Parents
I7:Family: Youth
I8:School: School
I9: School: Youth
I10:Job-seeking
I11:Leg. Sys. Inv.
I12:Contracting Act.
I13:Rel. Activities
I14:Advocacy Act.

Behavioral Measure

Observer Raters

R1:V. Soc. Comparison
R2:Resourceful.
R3:Judgmental
R4:Warmth
R5:Historical
R6:Empathy

Self

S1:V.Soc.Comparison
S2:Resourceful.
S3:Judgmental.
S4:Warmth
S5:Historical
S6:Empathy

Overall evaluation

OVR: observer raters
PR1:participant rater 1
PR2:participant rater 2

Table 31

Intercorrelations of Behavioral Measure Component Scores
with Intervention Scale Scores, Time 3

	R1	R2	R3	R4	R5	R6	S1	S2	S3	S4	S5	S6	OVR	PR1	PR2
I1	-.10	-.09	.08	-.11	.03	-.16	-.12	.01	-.11	.16	-.32	.08	-.05	-.15	-.01
I2	-.06	.03	.05	-.14	-.08	-.11	.01	.21	.02	.15	-.22	-.17	.10	-.24	-.02
I3	-.13	.17	-.03	.22	-.11	-.30	.10	.26	.16	.00	.06	-.16	.36	.08	.04
I4	-.08	.01	-.06	.19	-.19	-.17	.14	-.01	-.16	.14	.24	-.02	.12	.16	.06
I5	.18	-.01	.29	.05	-.08	-.06	.17	.14	.09	.13	-.13	-.05	-.05	-.18	-.01
I6	-.05	.17	-.12	.30	-.03	-.13	.22	.27	.01	.03	.37	.23	.47	.09	.12
I7	.08	.33	.00	.37	-.09	-.08	.29	.43	.02	.07	.33	.32	.51	.06	.08
I8	.04	.25	-.01	.03	-.15	-.13	.11	.27	.08	.18	.02	.21	.14	.02	-.08
I9	.14	.42	.14	.17	-.11	-.05	.22	.39	.07	.25	.29	.19	.31	-.06	-.07
I10	.02	.06	.09	-.02	-.14	-.04	.24	.16	.10	.01	.17	.19	.12	-.18	-.08
I11	.03	.12	.00	.05	-.09	.09	.17	.17	.00	.19	.01	.10	.08	-.03	-.05
I12	-.06	.15	-.04	.06	-.01	-.19	-.01	.18	.20	.00	.12	.17	.22	-.17	.02
I13	-.02	.27	-.03	.19	-.10	.15	.12	.28	-.07	.12	.24	.32	.47	.03	.12
I14	-.05	.23	-.05	.17	-.16	.04	.33	.34	.04	.18	.17	.32	.33	-.07	.04

Key:

Intervention Scale

I1:Vol./large Inv.
I2:Lack of Complaints
I3:Parental Inv.
I4:Peer Inv.
I5:Rec. Activities
I6:Family: Parents
I7:Family: Youth
I8:School: School
I9:School: Youth
I10:Job-seeking
I11:Leg. Sys. Inv.
I12:Contracting Act.
I13:Rel. Activities
I14:Advocacy Act.

Behavioral Measure

Observer Raters

R1:V.Soc.Comparison
R2:Resourceful.
R3:Judgmental
R4:Warmth
R5:Historical
R6:Empathy

Self

S1:V.Soc.Comparison
S2:Resourceful.
S3:Judgmental.
S4:Warmth
S5:Historical
S6:Empathy

Overall evaluation

OVR:observer raters
PR1:participant rater 1
PR2:participant rater 2

also focused more on changing both the school and the youth in school, were more involved in the legal system and engaged in more advocacy activities. Finally, those rated more highly on the Overall Evaluation by observer raters were more involved with the parents of their youth, focused more on changing the youth within their families, focused more on changing both school and the youth within the school context and engaged in more relationship and advocacy activities.

At time 2, there were 19 significant correlations of 210 correlations examined (Table 30). Students rated as more resourceful by the observer raters also focused more on changing the youth within school and were engaged in more relationship activities. Those rated more warm by observer raters tended to focus more on changing the youth within the family. Students who reported that they voiced more dissimilarities and similarities with the youth also focused more on changing the youth within the school context and engaged in more advocacy activities. Students reporting themselves as more resourceful focused on changing the youth both within the family and the school contexts more and engaged in more advocacy activities. If they rated themselves more highly on asking questions of the youth's past, they also tended to focus on the changing of both the parents and the youth within the family context. Those rating themselves as more empathic focused on changing the youth within the family more. A higher Overall Evaluation by the observer raters tended to go with those students who were more involved with the parents, focused on changing both the youth and

the parents in the family, focused on changing the youth in school, and engaged in more job-seeking, relationship activities and advocacy activities. A higher rating by the first participant rater indicated greater complaining and less positive involvement with their youth.

At time 3, there were 25 significant correlations of the 210 correlations (Table 31). Students rated as more resourceful by observer raters focused more on changing the youth in the family and the school settings. People rated as more judgmental by the observer raters tended to engage in more recreational activities. A higher rating of warmth by the observer raters was correlated with greater focus on changing both parents and youth within the family. Those rated more empathic by observer raters also were less involved with the youth's parents. Students who rated themselves as having voiced more comparisons with the adolescent in the behavioral measure focused more on changing the youth in the family, and participated in more advocacy activities. If the students rated themselves more highly on resourcefulness, they focused more on changing the youth both within the family and the school, as well as engaging in more relationship and advocacy activities. Students who rated themselves as more historical were less positively involved with and complained more about their youth, focused on changing the parents and the youth in the family and on changing the youth in school more. Students who reported that they labeled more feelings of the youth focused more on changing the youth in the family and engaged in more relationship and

advocacy activities. Finally, a high rating on the Overall Eavluation by the observer raters tended to point out those who were more involved in the youth's parents,focused more on changing both the parents and the youth in the family, focused more on changing the youth in school, and engaged in more relationships and advocacy activities.

In summary, it would appear that three of the behavioral measure factor scores were most consistently correlated (i.e., at a minimum of two of the three time periods) with certain Intervention scale scores. The self-rating of Resourcefulness was most consistently found correlated with Family: Focus on Changing Youth, School: Focus on Changing Youth, Relationship and Advocacy Activities. Historicalness as rated by self was most repeatedly associated with Family: Focus on Changing Parents, Family: Focus on Changing Youth, and School: Focus on Changing Youth. The Overall Evaluation score by the observer raters was consistently correlated with Parental Involvement, Family: Focus on Changing Parents, Family: Focus on Changing Youth, School: Focus on Changing Youth, Relationship Activities and Advocacy Activities. However, given the small number of students in the sample and the large number of correlations, extreme caution should be exercised in interpreting any of these relationships.

Relationships between overall in-class evaluation rank and PRF. There was one of four significant intercorrelations of the PRF factor scores and the overall in-class evaluation rank.

Outgoingness (PRF) was correlated with the in-class evaluation rank (.32). Thus, people who were more outgoing and less autnomous were ranked more highly be peers, supervisors and themselves on in-class evaluations.

Relationships between overall in-class evaluation rank with Intervention scale. There were no significant intercorrelations between the in-class evaluation rank score and any of the 14 Intervention scale scores at any of the three time periods. Thus, out of 42 possible correlations, none were significant.

Relationships between PRF and Intervention scale. Of the possible 168 correlations, there was only one significant inter-correlation of a PRF factor score and any of the 14 Intervention scale scores at any of the three time periods. Achievement-motivation, as measured by the PRF, was significantly correlated with Job-seeking from the Intervention scale at time one (.27). Those students who were more motivated and achievement-oriented tended to be involved in more employment-seeking activities with their youth at Time 1.

Summary. Of the 919 examined intercorrelations among the independent measures of this research, there were only 72 significant intercorrelations. This is slightly higher than that expected by chance alone; however, due to the intercorrelations within type of independent variable (especially the behavioral measure and the

and the Intervention scale scores), all of these relationships must be interpreted with great care and conservativeness.

Discriminant Analyses

The effects of training conditions upon the behavioral measure factor scores and the relationships of all of the independent measures have been reported in the preceding two sections. In this section, the results of the major design of the research will be reported. In this design, the independent measures were combined in a series of five step-wise discriminant function analyses in order to attempt to create a selection battery for the college students which could significantly predict those volunteers whose assigned youth were less likely to have further contact with the juvenile justice system.

Two-step factorization of predictors. Due to the large number of predictors (15 behavioral measure component scores, 42 Intervention scale scores, four PRF component scores and one overall in-class evaluation rank), it was decided to attempt a two-step factorization of the predictors in an attempt to further collapse predictors across types of measure (Golding and Seidman, 1974). However, due to the large degree of method variance, this second factorization using principal components analysis with varimax rotation merely produced another set of components within types of measures. In order to collapse the measures and still maintain a reasonable proportion of the variance, another tactic was employed.

Preliminary discriminant analyses. Three separate step-wise discriminant function analyses were conducted using the success-failure designation as the criterion. The Intervention scale scores of each time period were utilized as predictors; predictors were entered into the analysis only to the point at which there ceased to be a significant change in Rao's V at the $p < .05$ level (Nie, Hull, Jenkins, Steinbrenner & Bent, 1975). These three separate analyses yielded 10 significant predictors: Volunteer/Target Involvement (time 1); Family: Focus on Changing Parents, Family: Focus on Changing Youth, School: Focus on Changing Youth and Legal System Involvement (time 2); Recreational Activity, Family: Focus on Changing Parents, Family: Focus on Changing Youth, School: Focus on Changing Youth and Legal System Involvement (time 3).

These 10 predictors were then entered as predictors into a fourth step-wise discriminant function analysis, adding the 15 component scores of the behavioral measure as additional predictors of the success criterion. From this fourth analysis, there emerged a new combination of eight predictors. As seen in Table 32, six predictors remained from the original 10 entered from the Intervention scale scores; two component scores from the 15 of the behavioral measure also emerged as significant predictors as well. (See Table 32.)

Final discriminant analysis. In addition to the eight predictors generated from the fourth discriminant analysis, it was

Table 32

Variables Used as Predictors of Success
Final Discriminant Function Analysis

Intervention Scale Scores

Volunteer/Target Involvement, time 1
Family: Focus on Changing Parents, time 2
Family: Focus on Changing Youth, time 2
School: Focus on Changing Youth, time 2
Legal System Involvement, time 2
Legal System Involvement, time 3

Behavioral Measure Component Scores

Judgmentalness, rated by observer raters
Warmth, rated by self

Other Component Scores of Interest

Impulsivity (PRF)
Overall in-class evaluation rank
Training condition (4 dichotomous variables)

---

decided to include six other predictors in the final discriminant analysis, due to research interest. The PRF component score most highly correlated with success was included because of the common practice of administering paper and pencil measure to applicants for volunteer positions as a screening procedure. Since Impulsivity as measured by the PRF was most highly correlated with success (-.23), it was selected. (See Appendix A, Table A-4, for correlations of all measures with success.) The overall in-class evaluation rank was included to see if in-class skills was at all related to success. Finally, the four training conditions were entered as additional predictors in order to discern if they were in any way interacting with the success criterion. Thus, as shown in Table 32, in the final discriminant function analysis there were eight predictors from the fourth discriminant function analysis as well as the two new factor scores (Impulsivity, PRF and overall in-class evaluation rank) and four dichotomous measures of training condition.

Results of the final discriminant function analysis are reported in Table 33. All six of the Intervention scale scores were retained as predictors, as well as the measure of self-rated Warmth. According to this analysis, therefore, a "successful" volunteer had the following characteristics. S/he was lower on Volunteer/Target Involvement at time 1 (s/he spent less time with her/his youth); s/he was lower on both Family: Focus on Changing Youth and School: Focus on Changing Youth (s/he focused less on changing the youth within the context of family and school) at

Table 33

Final Discriminant Function Analysis

Standardsized Discriminant Coefficients

Variable	Coefficient	Group Centroids
Volunteer/Target Involvement, Time 1	-.64197	Success .71272
Family: Focus on Parents, Time 2	-.54640	Failure -2.59169
Family: Focus on Youth, Time 2	-.60422	
School: Focus on Youth, Time 2	1.37603	
Legal System Involvement, Time 2	-.60266	
Legal System Involvement, Time 3	-.73789	
Self-rated Warmth (behavioral measure)	.33151	

Prediction Results

	Predicted Group	
Actual Group	Success	Failure
Success	39	1
Failure	1	10

% Correctly Classified = 96%

$\chi^2 = 43.314$, $p < .01$ Canonical r = .81107

time 2; s/he scored higher on Family: Focus on Changing the parents at time 2 (s/he focused more on changing the parents); s/he scored lower on Legal System Involvement (s/he was less involved in the legal system with her/his youth) at both times 2 and 3; and s/he rated her/himself more warmly on the behavioral measure (s/he believed that s/he had appeared warmer, lectured the youth less and asked more open-ended questions during the situations of the behavioral measure).

Shrinkage. From this final discriminant function analysis, a canonical correlation coefficient of .8111 was calculated. Applying a formula for an estimate of the cross-validated multiple correlation (Darlington, 1968), $\hat{\rho}_c$ was computed to be .72. Of course, this correlation is inflated due to the methods employed, and should be interpreted with appropriate catuion.

CHAPTER IV

DISCUSSION

Prior to discussion of the results of this research and their implications, it is important to reiterate some crucial points. First, this study is based on a small number of subjects, especially given the number of variables involved in the study. Therefore, the power and the durability of the results may suffer. Similarly, the study is clearly meant to be one of discovery of possible relationships rather than one of supporting hypothesized relationships. Thus, particularly when dealing with the final discriminant analyses, some of the procedures may have violated statistical principles, however, with appropriate caution, some interesting issues and ideas are generated from this study. This section will first address the three major research questions posed in earlier sections; it will then propose a conceptual model placing the research in a meaningful context; finally, it will present some implications and conclusions which may be drawn from this study.

Research Questions Posed by This Study

This subsection will deal with the three major research questions posed in this study. First, the potential of the behavioral measure as a measure of training received will be

examined. Secondly, intercorrelations of all measures and predictive validity of the behavioral measure will be discussed. Finally, the "selection" battery generated will be examined.

Behavorial measure as a training outcome measure. The effects of the different training conditions on performance in the behavioral measure were examined. Clearly, the behavioral measure was able to detect training differences in students' styles of interacting with the youth. The people trained in the Action condition thought of more options and steps for their plans for the youth. This training condition was based upon the idea that the youth was placed in legal jeopardy due to two deficits: there was a lack of clear communication between the youth and those holding her/his reinforcers; and the youth suffered from unmet needs as well as problem areas which could be dealt with by generating and activating resources available in the environment to deal with these areas (Davidson and Rapp, 1976; Seidman and Rappaport, 1974; Stuart, 1971).

Those receiving the Relationship training labeled more of the youths' feelings and believed themselves to be voicing fewer comparisons between themselves and the youth while offering fewer opinions. In this training condition, students were trained that the youth had been placed in legal jeopardy due to her/his inability to deal with feelings involved with the problems and therefore an inability to solve problems on her/his own; it was therefore stressed that students should focus upon goals of

the youth, help the youth to explore her/his feelings concerning these goals, and to move on to the generation of options and the youth's implementation of her/his goals regardless of the student's own beliefs of correct ways to proceed (Goodman, 1972; Gordon, 1970' Rogers, 1961).

Students in the Low Intensity--Large Group were the most judgmental of all of the students. In this condition, students met once a month, in a group of 15, and, in essence, exchanged stories of their experiences with their youth. They were not offered any structured theory upon which to base their efforts, and often in effect received little guidance and few alternatives in working with their youth. Thus, students did not learn any new ways of viewing adolescents and their problems, and appeared to react in traditional ways of judging the adolescents in the behavioral measure and responding in opinionated ways to their problems. Thus, based on all of these findings, the behavioral measure not only was able to detect these differences, but people also acted in this setting as they had been trained (or not trained, as the case may be) to act.

Intercorrelations and predictive validity of the behavioral measure. The second area of interest examined was that of the interrelationships of the independent measures, with an eye toward especially the behavioral measure. It was discovered that there was a great deal of method variance in these measures. Except for the behavioral measure and the Intervention scale, there were no

meaningful correlations between types of measures. In addition, the behavioral measure did not correlate to a greater extent with the scale scores of the Intervention scale at the first time period than with the second or third time periods. However, one would expect that due to the closer proximity of the behavioral measure and the first time period of the Intervention scale, there would be more connections between how the students acted in the behavioral measure and how they acted with their assigned youth at the earliest time. This could imply either that the student was not acting the same way in the field or that the Intervention scale was not picking up the differences. A third possibility is that the students reacted to the behavioral measure as a kind of examination on what they had been trained to do, even though they were asked to act naturally; however, in the field working with their youth, they acted differently. Still a fourth possibility is simply that the behavioral measure has low concurrent intercorrelations and predictive validity: it does not correlate with other measures and it does not correlate significantly with the criterion. This possibility, however, seems an unlikely explanation, due to the general overall method variance which was discovered (none of the independent measures coverged with any of the others to any significant degree) and the complexity of the total conceptual model (to be delineated in greater detail in later subsections), direct intercorrelations and correlations with the criterion do not seem essential for considerations within the selection battery. For example, some of the Intervention scale

scores also did not significantly relate to success, but were significant predictors of success when coupled with other scores (e.g., Family: Focus on Changing Youth). A fifth explanation could be that the conception of the model placing the research in context is different from the linear model underlying this study--student assessment leading directly to target outcome. This possibility will be discussed in more detail also in a later subsection.

Potential selection battery. The third question examined by this research was that of a potential selection battery. There are three related issues involved in this question. First, in order to have a pure "selection" battery, only the behavioral measure and the PRF component scores could have bee included, for they were the only data collected prior to the students' involvement in the Project. In order to look at this question, a discriminant analysis was computed using only the PRF component scores and the behavioral measure scores; however, no scores entered the prediction equation with any significance. (A similar result was obtained utilizing only PRF component scores, behavioral measure component scores and the training conditions as predictors.) Therefore, when taken alone, none of the data collected at the pre time period were able to predict success. However, when taken with the Intervention scale scores, still only one score (self-rated Warmth from the behavioral measure) added significantly to the prediction capabilities.

Lack of predictive results of the pre measures is a second issue. It is not especially surprising that the paper and pencil measure (PRF) did not predict well, for other researchers have reported similar findings (e.g., Mordock and Platt, 1969; Rappaport, 1977; Rappaport, Chinsky and Cowen, 1971). However, since the behavioral measure was constructed with an eye toward its predictive capabilities, it is important to examine its lack of prediction.

Of the studies reviewed in this paper, the most frequently used behavioral measure in mental health studies was the GAIT (e.g., Dooley, 1975; Goodman, 1971). This measure approached the criterion situation and showed slight intercorrelations. For example, there were some intercorrelations between peer and observer raters (D'Augelli, Chinsky and Getter, 1974; Rappaport, et al., 1971). However, in the GAIT there was only one situation presented; there was therefore no attempt to even look at generalizability across situation. Dooley (1975) even reported trouble with generalizability across time (a nine-week interval). In addition to this, predictive validity of the GAIT was very low (Goodman, 1972; Rappaport, et al., 1971). Thus, the only available behavioral measure was fraught with numerous difficulties and unaddressed issues.

Looking at the new behavioral measure created in this study, there was little generalizability across situations; only the self-ratings showed that the situations were the same, there was even little generalizability across time (as little as 30

minutes). There was also little generalizability across type of raters. Therefore, it is not particularly surprising, statistically, that the measure did not predict success.

But why was there such a lack of generalizability? In this research, a measure was created which was as close to the criterion situation as realistically possible. There were adolescents, interacting in person with the students; the adolescents as well as trained observer raters rated the students; the multiple situations about which the students and the adolescents interacted were carefully selected from very typically occurring problems between adolescents and college students in the Project. In addition to all of this, the observer raters were carefully trained in order to increase the reliability of the instrument. And yet, in spite of these optimal conditions, there still was no ability to predict success. (It must of course be kept in mind that the group of 57 students measured in the behavioral measure was already a very select group, due to the involved self-selection process involved in the Project. Thus, there is always the possibility that there simply was too little variance in the students' behavior--except in the few categories where the students were trained or not trained with divergent goals--to be picked up by the measure at this point.) Thus, one conclusion which could be drawn is that even under the most realistically optimal conditions possible for a selection measure--a live observation under circumstances as close to the working situation as possible without

employing the applicant--it was still impossible to predict a successful volunteer from pre data.

A third issue pertaining to the selection battery generated here concerns the inclusion of the Legal System Involvement scale scores in the battery. It could be claimed that by doing so, prediction of success was ensured, since a variation of the criterion was included in the battery. Indeed, the Legal System Involvement scale scores for the three time periods alone as predictors correctly classified 83% of the volunteers. (Considering the base rate alone, predicting everyone to be a success could have correctly classified 68% of the volunteers.) However, the addition of the other scale scores added some additional information about the Intervention process which was lacking from considering only the legal system involvement.

Conceptual and Methodological Findings

There are three important sets of issues that this research brings to light. First, a conceptual model giving references to the research in this study will be presented. Next, questions related to this model will be examined. Thirdly, findings involving the failure of the participant raters to produce any significant and meaningful measures will be discussed.

Conceptual model. As an overview of the context within which this research was conducted, Figure 3 represents the various aspects of the research and their interrelationships. Looking first

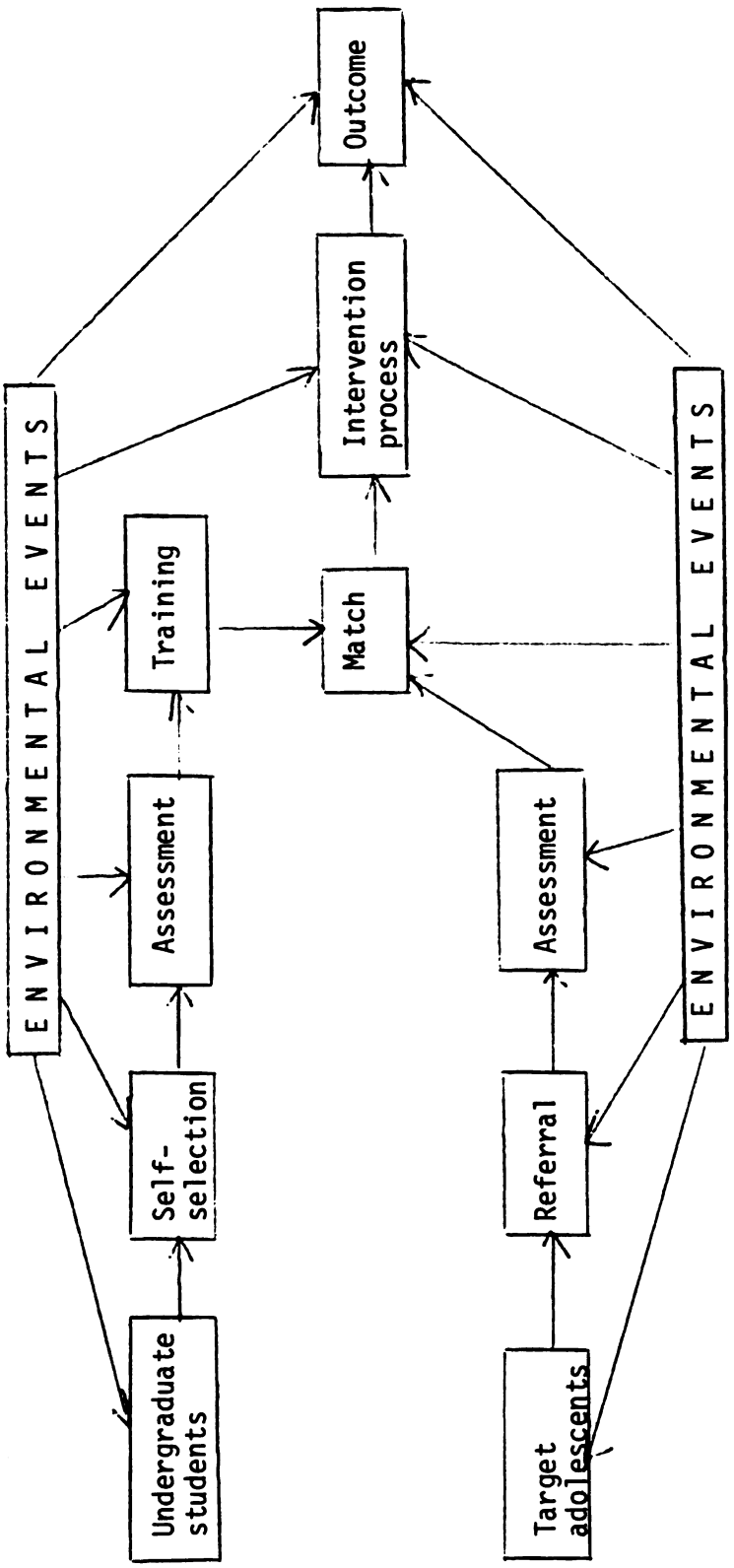


Figure 3. Conceptual Model

at the target youth, there was a pool of adolescents in the Probate Court (TARGET YOUTH) who were referred to the Project (REFERRAL); they were assessed by the Project and were thus readied for matching with a student (ASSESSMENT). At the same time, there was a pool of undergraduate students (from a larger pool of undergraduate Social Science majors--UNDERGRADUATE STUDENTS). These 134 students were assessed by the Project and randomly selected (ASSESSMENT). The 57 randomly selected were then trained and also "readied" for work with the youth (TRAINING). Then, a student and youth were matched (MATCH), and intervention ensued for 18 weeks (INTERVENTION PROCESS) and an outcome was measured (OUTCOME). However, as illustrated in Figure 3, environmental events were constantly at work on both the student and the youth as well as the match between the student and the youth. A multitude of factors entered into these events for both parties of the match, including such things as family and peer relations, attitudes, labeling processes, important life events were tapped through process interviews, by asking about such areas as active parental control, involvement in the school system and parental knowledge of friends. However, most of the points of the model were either partial or complete unknowns, such as the match of the student and the youth, the student's environmental events, and the majority of the processes involved--how the match interacted with the intervention process, how both the student and the target youth decided to be involved with the Project and how assessment affected the match and later

processes. Thus, Figure 3 presents a complex model of interaction within which this study was conducted.

Questions raised by the conceptual model. Specifically in the target youths' lives, there were numerous important sources of variance--for example, employment relations, peer relations, family relations, educational issues (e.g., Conger, 1976; Klein, Teilmann, Lincoln & Labin, 1978; Elliot & Voss, 1974). Similarly for the students, there were all of the above areas as well as variance due to training and supervision within the class (Goodman, 1971; Kantrowitz, Davidson, Blakely & Kushler, 1978). Further, those areas which have been monitored by the Project (primarily areas concerning the youths' environmental events) indeed have been found to be significantly related to success in the Project. Of 13 scales dealing with the youth's "life domain," six were significantly ($p < .05$) related to success: Parental Knowledge of Friends, Positive Change in School, Parental Impact on School, Involvement in School System, Involvement in the Juvenile Justice System, and Positive Change in Home. Of the 14 Intervention Scales, eight were significantly related to success ($p < .05$): Legal System Involvement; School: Focus on Changing School; School: Focus on Changing Youth; Contracting Activities; Volunteer/Target Involvement; Parental Involvement; Peer Involvement; and Relationship Activities. In addition to these separate sources of variance, there were the dynamics of the match between the student and the youth about which there was no information

collected in this design (Seidman & Rappaport, 1974). Thus, all of these areas contributed variance which was neither regulated nor monitored by the study referred to in this paper. Finally, it is very possible for the youth that the criterion as defined herein was much more socially defined rather than behaviorally defined. Of course, the youth's apprehension by the police or appearance before the Probate Court was a behavioral criterion; however, due to the labeling processes and heightened awareness by law officials concerning youth already in legal trouble, there were many other factors entering the definition of the criterion as well. Thus, the outcome criterion may have been only minimally related to what the youth actually did (Gold, 1970; Klein & Carter, 1976).

Thus, this research took place within the context of individual differences of students, individual differences of adolescents, differing environment events for both students and youth, and numerous unknowns entering into the match of student and youth, the intervention process and the labeling process involved in the definition of the outcome criterion, to name only a few variables. These variables are present in all naturalistic social science research as well; however, they are seldom acknowledged and dealt with in such settings. Therefore, due to this large number of uncontrolled and unacknowledged variables, it is evident that selection in such areas is destined to be unsuccessful.

Methodological issues surrounding participant raters. The third issues of importance here are raised concerning the participant raters. It was hoped that the adolescents' ratings of the undergraduates could provide a new and meaningful measure of the volunteers. After all, it would seem that no one is better qualified to judge the potential of a volunteer than one who had had interactions of a similar kind with another volunteer. This, however, proved not to be the case at all. The adolescents' rating on both the individual characteristics and the overall evaluation neither converged nor were consistently related to success of the volunteer in the long run. Part of the difficulty was found in the lack of variance of the youths' ratings: 21 of the 22 ratings had at least 80% of the responses in two contiguous response categories. This could have occurred particularly in this study because of the use of tally sheets for the observer raters, and the training of both observer and participant raters to make frequency counts. However, the participant raters had extreme difficulty in keeping track of their counts. Even though they rated each person within one minute of the end of the interaction, it was very difficult for them to remember counts of 20 different characteristics.

Why were these raters so different from the other raters in both this study and other studies? (It should be remembered that even peer raters in the GAIT measure had more convergence than the participant raters in this behavioral measure.) First, the adolescents were not college undergraduates interested in human

service careers as were the other raters of this measure and the raters of the GAIT. Therefore, in this respect, it is not unexpected that they would rate people differently. Also, the types of ratings on the GAIT were substantially different from those in this behavioral measure: the GAIT dealt mainly with overall therapeutic characteristics which people interested in human service are generally at least vaguely familiar with. The behavioral measure created here dealt with a few therapeutic characteristics and also with specific, countable behaviors which needed to be enumerated by the raters and both of which the adolescents may have not been accustomed to paying attention to (e.g., empathic, number of plans generated). And of course, there is always the issue of which type of rater is "really" capturing the interaction best. Simply because the adolescents do not converge with other types of people does not necessitate their exclusion from consideration.

It is possible that by drilling the adolescents alone on typical responses and having them categorize the responses, greater variance could have resulted from the participant rater scores. As it was, it appears that the adolescents did not differentiate much from one volunteer's responses to the next. Another possibility would be the creation of some kind of unobtrusive counter to enable the youth to count some of the more frequently occurring responses (e.g., open-ended and close-ended questions, number of plans generated).

Implications and Conclusions

Again, it is important to re-emphasize that there was a small sample size involved in this research. It is possible that when there is a larger group of students to examine (as will occur as the Project continues), findings might be different from those reported here. However, with this caution well in hand, the following are the conclusions which can be drawn from this study.

First, the behavioral measure is a relatively accurate method of determining training differences. This, however, does not predict the success of the student and the youth during their intervention period. A much less expensive and more reliable method of determining success is to ask the people involved in the process what is being done. Thus, as a training measure, the behavioral measure may be far too expensive, considering the potential gains from it.

Secondly, the utility of the behavioral measure as a method of selecting students who will be successful appears to be very limited. It is quite expensive to administer: for the Project (with inexpensive student labor and subsidized space and supplies), the cost was approximately $15 per student. For an outside agency, this figure could easily rise to $75 or $100 per student. This expense is clearly not outweighed by a higher utility than other more simple and inexpensive measures; if this were the case, there might be cause for still utilizing a high utility-high expense measure (Wiggins, 1973). Also, conceptually, the behavioral measure does not plug into the model presented in

Figure 3 very adequately. Although there are may processes involved in the model, the design of this research (and most selection studies in mental health, for that matter) is purely linear in nature: nonprofessional assessment ——> outcome. As evidenced by this research, this model is not an adequate one to predict outcome of this nature.

Selection studies often focus on paper and pencil measures to tap into personality traits. These have been found to be inadequate in predicting success. The behavioral measures which have been utilized were found to be lacking in predictive validity. Based upon theories of prediction from testing, a measure was created which offered four situations rather than merely the one offered in the GAIT within which to view the volunteer (Mischel, 1968); raters were trained in behavioral ratings as opposed to therapeutic hunches (Guilford, 1959); and the situations observed were made more face valid than those in other behavioral measures (Rotter, 1960; Durlak, 1971). In spite of these changes, the measure was still unable to predict the success of the volunteer in the Project.

This brings about consideration of the issue of selection in mental health programs. There is a plethora of literature attempting to find sound selection methods for nonprofessionals (e.g., Goodman, 1972; Rappaport, et al., 1971; Stollak, 1968; Suinn, 1974a & b). However, reflecting on the conceptual model proposed in Fgiure 3, it becomes apparent why no one has been successful in discovering valid selection techniques. There are

so many unknowns--either unmonitored or currently unmonitorable--that as the state of the art stands at this moment, it is impossible to track that which is assessed in the selection process through to the assessment of the final outcome. Thus, it would appear that the search for a selection method could be temporarily skirted and efforts might more rightfully be focused upon delineating in more specificable terms that which is occuring at each of the steps of the model in order to better understand the workings of the model. Until these components can be more clearly understood, the issue of selection methods is at best an academic one.

These issues, however, can be applied even more generally than merely to the mental health field. As shown in Figure 3, in naturalistic settings, there is a myriad of variables which needs to be scrutinized. Too little attention is paid to the processes involved in any long-term interaction and all of the individual differences which can come into play within such a system--targets, interveners, environmental events, supervisors, data gathers, etc. In this model of interaction, all of these can have impact upon the final outcome observed and measured. In general, therefore, this process needs to be more succinctly investigated before social science researchers can hope to make impact not only upon methods of selecting those who will enter and operating within a system successfully, but make impact upon the system at any chosen vantage point--be that assessment,

selection, intervention, mathcing or any of the processes occurring. Without this knowledge, researchers cannot expect to fully grasp the holistic functioning of social systems.

REFERENCES

REFERENCES

Aiken, J., Brownell, A., & Iscoe, I. The training and utilization of paraprofessionals in a college psychological service center. Journal of College Student Personnel, 1974, 15, 480-496.

Albee, G. W. Conceptual models and manpower requirements in psychology. American Psychologist, 1968, 23, 317-320.

Balma, M. J. The concept of synthetic validity. Personnel Psychology, 1959, 12, 395-396.

Bass, A. R. Personnel selection and evaluation. In S. E. Seashore & R. J. McNeill (Eds.), Management of the urban crisis. New York: Free Press, 1971.

Bohr, V. C. The use of students as employees in a community mental health center. In A. Biegel & A. I. Levenson (Eds.), The community mental health center: Strategies and programs. New York: Basic Books, Inc., 1972.

Brown, W. F. Student-to-student counseling for academic adjustment. Personnel and Guidance Journal, 1965, 43, 811-817.

Brown, W. F. Effectiveness of paraprofessionals: The evidence. Personnel and Guidance Journal, 1974, 53, 257-263.

Brown, W. F., Wehe, N. O., Zunker, V. G., & Haslom, W. L. Effectiveness of student-to-student counseling on the academic adjustment of potential college dropouts. Journal of Educational Psychology, 1971, 62, 285-289.

Buckley, H. M., Muench, G. A., & Sjoberg, B. M. Effects of college student visitation program on a group of chronic schizophrenics. Journal of Abnormal Psychology, 1970, 75, 242-244.

Carkhuff, R. R. Helping and human relations (Vol. 1): Selection and training. Holt, Rinehart and Winston, Inc., 1969.

Chinsky, J. M. Nonprofessionals in mental health (Doctoral dissertation, University of Rochester, 1968). Dissertation Abstracts, 1969, 30, 1355B. (University Microfilms No. 69-14, 769).

Chinsky, J. M., & Rappaport, J. Evaluation of a technique for the behavioral assessment of nonprofessionals. Journal of Clinical Psychology, 1971, 27, 400-402.

Codori, C., & Cowles, J. T. The problem of selecting adults for a child care training program: A descriptive and methodological study. Child Care Quarterly, 1971, 1, 47-55.

Conger, R. D. Social control and social learning models of delinquent behavior: A synthesis. Criminology, 1976, 14, 17-39.

Cowen, E. L., Zax, M., & Laird, J. D. A college student volunteer program in the elementary school setting. Community Mental Health Journal, 1966, 2, 319-328.

Cronbach, L. J. Essentials of psychological testing. New York: Harper and Row, 1970.

Cronbach, L. J., & Gleser, G. C. Psychological tests and personnel decisions (2nd ed.). Urbana, Il.: University of Illinois Press, 1965.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, H. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: John Wiley & Sons, 1972.

Darlington, R. B. Multiple regression in psychological research and practice. Psychological Bulletin, 1968, 69, 161-182.

D'Augelli, A. R., Chinsky, J. M., & Getter, H. The effects of group composition and duration on sensitivity training. Small Group Behavior, 1974, 5, 56-64.

Davidson, W. S., II, & Rapp, C. A multiple strategy model of child advocacy. Social Work, 1976, 21, 225-232.

Davidson, W. S., II, Seidman, E., Rappaport, J., Berck, P. L., Rapp, N. A., Rhodes, W., & Herring, J. Diversion program for juvenile offenders. Social Work Research and Abstracts, 1977, 13, 41-49.

Davison, G. C. The training of undergraduates as social reinforcers for autistic children. In L. Ulmann & L. Krasner (Eds.), Case studies in behavioral modification. New York: Holt, 1965.

Dawson, R. M. Personality and peer counseling: An Australian study. Personnel and Guidance Journal, 1973, 52, 46-49.

Delworth, U., & Moore, M. Helper plus trainer: A two-phase program for the counselor. Personnel and Guidance Journal, 1972, 52, 428-433.

Dooley, D. Assessing nonprofessional mental health workers with the GAIT: An evaluation fo peer ratings. American Journal of Community Psychology, 1975, 3, 99-110.

Dorr, D., Cowen, E. L., Sandler, I., & Pratt, D. M. Dimensionality of a test battery for nonprofessional mental health workers. Journal of Consulting and Clinical Psychology, 1973, 41, 181-185.

Durlak, J. The use of nonprofessionals as therapeutic agents: Research, issues and implications. Unpublished doctoral dissertation, Vanderbilt University, 1971.

Elliot, D. S., & Voss, H. L. Delinquency and dropout. Lexington, Ma.: Lexington Books, 1974.

Empey, L. T. Alternatives to institutionalization. Washington, D.C.: Government Printing Office, 1967.

Erickson, M. L. Group violations, socioeconomic status and official delinquency. Social Forces, 1973, 52, 41-52.

Federal Bureau of Investigation, Uniform crime reports. Washington, D.C.: Government Printing Office, 1973.

In re Gault, 387U.S. 1 (1967).

Gold, M. Delinquent behavior in an American city. Belmont, Ca.: Brooks-Cole, 1970.

Golding, S., & Seidman, E. Analysis of multitrait-multimethod matrices: A two-step principal components procedure. Multivariate Behavioral Research, 1974, 9, 479-496.

Goodman, G. Companionship therapy. San Francisco: Jossey-Bass, 1972.

Gordon, T. Parent effectiveness training. New York: Peter H. Wyden, 1970.

Gruver, G. G. College students as therapeutic agents. Psychological Bulletin, 1971, 76, 111-127.

Guilford, J. P. Fundamental statistics in psychology and education. New York: McGraw-Hill Book Co., Inc., 1942.

Hauer, A. L. Adjunct counselors in college. Personnel and and Guidance Journal, 1973, 52, 43-46.

Holzberg, J. D., Knapp, R. H., & Turner, J. L. College students as companions to the mentally ill. In E. L. Cown, E. A. Gardner, & M. Zax (Eds.), Emergent approaches to mental health problems. New York: Appleton-Century-Crofts, 1967.

Irwin, J. The trouble with rehabilitation. Criminal Justice and Behavior, 1974, 1, 139-149.

Jackson, D. N. Personality Research Form manual. Goshen, N.Y.: Research Psychologists Press, Inc., 1974.

Jackson, D. N. A sequential system for personality scale development. In C. Spielberger (Ed.) Current topics in clinical and community psychology (Vol. 2). New York:

Johnson, C. Paraprofessionals briding the gap. American Journal of Orthopsychiatry, 1971, 41, 234-235.

Kalafat, J., & Tyler, M. The community approach: Programs and implications for a campus mental health agency. Professional Psychology, 1973, 4, 43-49.

Kaiser, H. F., & Caffry, J. Alpha factor analysis. Psychometrika, 1965, 30, 1-14.

Kantrowitz, R. E., Davidson, W. S., II, Blakely, C. H., & Kushler, M. G. The effect of training/supervision of nonprofessional interventions with delinquents. Paper presented at the Annual Convention of the American Psychological Association.

Karlsruher, A. E. The nonprofessional as a psychotherapeutic agent: A review of the empirical evidence pertaining to his effectiveness. American Journal of Community Psychology, 1974, 2, 61-77.

Kelley, S. M., Shemberg, K. M., & Ferber, H. The training and use of undergraduates as behavior analysts in the consultative process. Professional Psychology, 1973, 4, 59-63.

King, L. W., & Turner, R. D. Teaching a profoundly retarded adult at home by nonprofessionals. Journal of Behavior Therapy and Experimental Psychiatry, 1975, 6, 117-121.

Klein, M., & Carter, R. Off of the streets: Diversion in the junvenile justice system. New York: Wiley, 1976.

Knapp, R. H., & Holzberg, J. D. Characteristics of college students volunteering for service to mental patients. Journal of Consulting Psychology, 1964, 28, 82-85.

Kopita, R. R. A description and evaluation of a program to train undergraduate student peer counselors to communicate understanding, acceptance and sincerity (Doctoral dissertation, University of Michigan, 1973). Dissertation Abstracts International, 1974, 34, 4746A-4747A. (University Microfilms No. 74-3664.)

Kreitzer, S. F. College students in a behavior therapy program with hospitalized emotionally disturbed children. In B. G. Guerney, Jr. (Ed.). Psychotherapeutic agents: New roles for nonprofessionals, parents and teachers. New York: Holt, Rinehart and Winston, 1969.

Lawton, M. P., & Lipton, M. B. Student employees become companions to patients. Mental Hospitals, 1963, 14, 550-556.

Levine, C. Impact of work with mental patients on student volunteers. Journal of Human Relations, 1966, 14, 422-433.

Levitt, E. L. Research on psychotherapy with children. In A. Bergin & S. L. Garfield (Eds.), Handbook on psychotherapy and behavior change. New York: Wiley, 1971.

Mischel, W. Personality and assessment. New York: John Wiley and Sons, Inc., 1968.

Mitchell, W. E. Amicatherapy: Theoretical perspectives and an example of practice. Community Mental Health Journal, 1966, 2, 307-314.

Mordock, J. B., & Platt, H. Selection and training of child-care workers. Hospital and Community Psychiatry, 1969, 20, 227-229.

Newton, F. B. The effect of systematic communication skills training on residence hall paraprofessionals. Journal of College Student Personnel, 1974, 15, 366-369.

Nie, N. H., Hull, C. H., Steinbrenner, K., & Bent, D. H. Statistical package for the social sciences (2nd Ed.). New York: McGraw-Hill Book Co., Inc., 1975.

Nietzel, M. F., Winett, R. A., MacDonald,M. L., & Davidson, W. S., II. Behavioral approaches to community psychology. New York: Pergammon Press, 1977.

Nunnally, J. Introduction to psychological measurement. New York: McGraw-Hill Inc., 1970.

Persons, R. W., Clar, C., Persons, M., Kadish, M., & Patterson, W. Training and employing undergraduates as therapists in a college counseling service. Professional Psychology, 1973, 4, 170-178.

Poser, E. G. The effects of therapists' training on group therapeutic outcome. Journal of Consulting Psychology, 1966, 30, 283-289.

Rappaport, J., Chinsky, J. M., & Cowen, E.L. Innovations in helping chronic patients: College student in a mental institution. New York: Academic Press, 1971.

Rappaport, J. Community psychology: Values, research and action. New York: Holt, Rinehart and Winston, 1977.

Roberts, A. G. The effect of systematic selection criteria and human resource development training on resident assistants' effectiveness within the residence hall environment (Doctoral dissertation, University of Tennessee, 1975). Dissertation Abstracts International, 1976, 36, 7216A-7217A. (University Microfilm No. 76-11, 084)

Rogers, C. On becoming a person. Boston: Houghton Mifflin, 1961.

Rotter, J. B. Some implications of a social learning theory for the prediction of goal directed behavior from testing procedures. Psycholgoical Review, 1960, 67, 301-316.

Scheibe, K. E. College students spend eight weeks in mental hospital: A case report. Psychotherapy: Theory, Research and Practice, 1965, 2, 117-120.

Schnelle, J. F., McNees, P., Huff, T. M., Marshal, R. S., & Hannah, J. F. Behavior intervention teams: The utilization of paraprofessionals in a community mental health center. Journal of Community Psychology, 1975, 3, 258-265.

Seidman, E., & Rappaport, J. The educational pyramid: A paradigm for training, research and manpower utilization in community psychology. American Journal of Community Psychology, 1974, 2, 119-130.

Seidman, E., Rappaport, J., Davidson, W. S., & Linney, J. Changing human service systems: Intervention with children, adolescents, adults and the elderly. New York: L. Erlbaum Assoc., in press.

Spoerl, O. H. An activity-centered volunteer program for university students. Hospital and Community Psychiatry, 1968, 19, 114-116.

Stollak, G. E. The experimental effects of training college students as play therapists. Psychotherapy: Theory, Research and Practice, 1968, 5, 77-80.

Stollak, G. E., Scholom, A., Green, L., Schreiber, J., & Messe, L. A. The process and outcome of play encounters between undergraduates and clinic-referred children: Preliminary findings. Psychotherapy: Theory, Research and Practice, 1975, 12, 327-331.

Stuart, R. B. Behavioral contracting within the families of delinquents. Journal of Behavior Therapy and Experimental Psychiatry, 1971, 2, 1-11.

Stuckey, B., Garrett, M. W., & Sugar, M. Group supervision of student companions to psychotic children. International Journal of Group Psychotherapy, 1971, 21, 301-309.

Suinn, R. M. Training undergraduate students as community behavior modifcation consultants. Journal of Counseling Psychology, 1974a, 21, 71-77.

Suinn, R. M. Traits for selection of paraprofessionals for behavior-modification consultation training. Community Mental Health Journal, 1974b, 10, 441-449.

Taylor, H. C., & Russell, J. T. The relationship of validity coefficients to the practical effectiveness of tests in selection. Journal of Applied Psychology, 1939, 23, 565-578.

Truax, C., & Carkhuff, R. Toward effective counseling and psychotherapy. Chicago: Aldine Publishing Co., 1967.

Umbarger, C. C., Dalsimer, J. S., Morrison, A. P., & Breggin, P. R. College students in a mental hospital. New York: Grune and Stratton, 1962.

Vesprani, G. J. Personality correlates of accurate empathy in a college companion program. Journal of Consulting and Clinical Psychology, 1969, 33, 722-727.

Walker, C. E., Wolpin, M., & Fellows, L. The use of high school and college students as therapists and researchers in a state mental hospital. Psychotherapy: Theory, Research and Practice, 1967, 4, 186-188.

Wasserman, C. W., McCarthy, B. W., & Ferree, E. H. Student para-professionals as behavior change agents. Professional Psychology, 1975, 6, 217-223.

Wiggins, J. S. Personality and prediction: Principles of personality assessment. Reading, Ma.: Addison-Wesley Publishing Co., 1973.

Zax, M., & Cowen, E. L. Early identification and prevention of emotional disturbance in a public school. In E. L. Cowen, E. A. Gardner, & M. Zax (Eds.), Emergent approaches to mental health. New York: Appleton-Century-Crofts, 1967.

Zax, M., & Specter, G. A. An introduction to community psychology. New York: John Wiley & Sons, 1974.

Zunker, V. G., & Brown, W. F. Comparative effectiveness of student and professional counselors. Personnel and Guidance Journal, 1966, 44, 738-744.

Reference Note

1. Lewis, R. G., & Davidson, W. S., II. Executive Summary: Model Evaluation Project: Youth Service Bureaus in Michigan. Lansing, Mi.: Office of Criminal Justice Programs, 1977.

APPENDICES

APPENDIX A

APPENDIX A

Table A-1

Intercorrelations within Participant
Rater by Characteristic
(Behavioral Measure)

	Rater 1, Situation 1/ Rater 1, Situation 2	Rater 2, Situation 3/ Rater 2, Situation 4
optimistic	.28	.30
pessimistic	-.03	.23
voiced similarities--now	.22	.08
voiced dissimilarities--now	.04	.05
voiced similarities--then	.19	*
voiced dissimilarities--then	*	*
historical	-.08	*
judgmental	.31	.30
resourceful	.07	-.08
specific planning	*	-.01
empathic	*	-.04
discounting	*	*
warm	-.03	.55
questioning style	.06	.18
impatience	.03	.31
talkative	.14	.06

*Correlation coefficients uncalculable due to lack of variance

Table A-2

Intercorrelations across Participant Rater and across Situations by Characteristic (Behavioral Measure)

Characteristic	Rater 1, Sit. 1/ Rater 2, Sit. 3	Rater 1, Sit. 1/ Rater 2, Sit. 4	Rater 1, Sit. 2/ Rater 2, Sit. 3	Rater 1, Sit. 2/ Rater 2, Sit. 4
optimistic	-.14	.03	-.01	.01
pessimistic	-.05	-.01	.00	.03
voiced similarities--now	.38	-.07	.08	-.12
voiced dissimilarities--now	.04	-.09	-.02	.05
voiced similarities--then	.27	*	-.14	*
voiced dissimilarities--then	*	*	.01	*
historical	.24	*	.10	*
judgmental	-.25	.08	-.09	.03
resourceful	.27	-.16	.16	.09
specific planning	*	*	-.01	.34
empathic	.31	-.04	*	.42
discounting	*	*	-.02	*
warm	-.06	.09	.13	-.05
questioning style	-.01	-.15	.11	-.10
impatience	-.05	-.01	-.54	.01
talkative	.01	.14	-.02	-.06

*Correlation coefficient uncalculable due to lack of variance

Table A-3

Intercorrelations of Observer Raters' Overall Characteristics
(Behavioral Measure)

	Rater 1 Char. 1	Rater 1 Char. 2	Rater 1 Char. 3	Rater 2 Char. 1	Rater 2 Char. 2	Rater 2 Char. 3	Rater 3 Char. 1	Rater 3 Char. 2	Rater 3 Char. 3
Rater 1, char. 1 (quality of options)									
Rater 1, Char. 2 (ease of interaction)	.65[a]								
Rater 1, Char. 3 (likelihood of effectiveness)	.77[2]	.66[a]							
Rater 2, Char. 1 (quality of options)	.50[b]	**	.50[b]						
Rater 2, Char. 2 (ease of interaction)	.50[b]	**	.50[b]	.47[d]					
Rater 2, Char. 3 (likelihood of effectiveness)	**	**	**	.86[d]	.51[d]				
Rater 3, Char. 1 (quality of options)	.69[c]	1.00[c]	.90[c]	.30[e]	1.00[e]	.30[e]			
Rater 3, Char. 2 (ease of interaction)	.35[c]	.90[c]	.64[c]	.30[e]	1.00[e]	.30[e]	.49[f]		

Table A-3 (Cont.)

	Rater 1 Char. 1	Rater 1 Char. 2	Rater 1 Char. 3	Rater 2 Char. 1	Rater 2 Char. 2	Rater 2 Char. 3	Rater 3 Char. 1	Rater 3 Char. 2	Rater 3 Char. 3
Rater 3, Char. 3 (likelihood of effectiveness)	.93[c]	.57[c]	.87[c]	.52[e]	.58[e]	.52[e]	.53[f]	.27[f]	

**Incalculable, due to lack of variance

[a]N = 23
[b]N = 3
[c]N = 4
[d]N = 24
[e]N = 4
[f]N = 20

Table A-4

Correlations of Independent Measures with Success

Component (Scale) Score	Correlation with Success
PRF	
Achievement-Motivation	-.00
Impulsivity	-.23
Outgoingness	-.01
Aggression	.07
Behavioral Measure	
Observer Raters	
Voiced Social Comparison	-.02
Resourcefulness	.14
Judgmentalness	.14
Warmth	-.01
Historicalness	-.16
Empathy	.07
Overall Evaluation, Rater 1	.20
Overall Evaluation, Rater 2	-.00
Overall Evaluation, Rater 3	.14
Self Raters	
Voiced Social Comparison	.05
Resourcefulness	.18
Judgmentalness	-.01
Warmth	-.05
Historicalness	-.03
Empathy	.07
Participant Raters	
Overall Evaluation, Rater 1	-.04
Overall Evaluation, Rater 2	-.04
In-Class Rankings	
Peer Ranking, time 1	.15
Supervisor Ranking, time 1	.18
Self Ranking, time 1	.20
Peer Ranking, time 2	.03
Supervisor Ranking, time 2	.18
Self Ranking, time 2	.27
Peer Ranking, time 3	.09
Supervisor Ranking, time 3	.27
Self Ranking, time 3	.27

Table A-4 (Cont.)

Component (Scale) Score	Correlation with Success
Intervention Scale	
Time 1	
Vol./Target Involvement	.39
Lack of Complaints/+ Involvement	.09
Parental Involvement	.09
Peer Involvement	.06
Recreational Activities	.10
Family: Focus on Parents	-.20
Family: Focus on Youth	.06
School: Focus on School	.14
School: Focus on Youth	.13
Job-seeking	-.11
Legal System Involvement	.01
Contracting Activities	.00
Relationship Activities	.02
Advocacy Activities	.06
Time 2	
Vol./Target Involvement	-.06
Lack of Complaints/+ Involvement	.13
Parental Involvement	.12
Peer Involvement	.12
Recreational Activities	-.03
Family: Focus on Parents	-.06
Family: Focus on Youth	.25
School: Focus on School	.29
School: Focus on Youth	.46
Job-seeking	-.10
Legal System Involvement	.46
Contracting Activities	.08
Relationship Activities	.04
Advocacy Activities	.10
Time 3	
Vol./Target Involvement	.01
Lack of Complaints/+ Involvement	.06
Parental Involvement	.10
Peer Involvement	.10
Recreational Activities	-.03
Family: Focus on Parents	-.09
Family: Focus on Youth	.18
School: Focus on School	.23
Job-seeking	-.07
Legal System Involvement	.48
Contracting Activities	.30
Relationship Activities	-.03
Advocacy Activities	.10