5-1981

Social validation: An investigation of factors that influence social validity ratings

Dianne J. Smart
University of Nebraska at Omaha

SOCIAL VALIDATION:

AN INVESTIGATION OF FACTORS

THAT INFLUENCE SOCIAL VALIDITY RATINGS

A Thesis

Presented to the

Department of Psychology

and the

Faculty of the Graduate College

University of Nebraska

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

University of Nebraska at Omaha

by

Dianne J. Smart

May, 1981

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

THESIS ACCEPTANCE

Accepted for the faculty of the Graduate College, University of Nebraska, in partial fulfillment of the requirements for the degree Master of Arts, University of Nebraska at Omaha.

Thesis Committee:	Name	Department
	Shelton Hendricks	Psychology
	Karen B. Maloney	Psychology
	Michael L. Hall	Educational Foundations

Chairman

April 22, 1981
Date

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank all the individuals who donated their time and energy to help make this thesis a reality. First, my sincerest appreciation goes to Dean Fixsen. The encouragement, expertise, support and concern Dean has given me will never be forgotten. I would also like to thank the other members of my committee --- Karen Blase, Shelton Hendricks and Michael Katz --- for their time, guidance and flexibility. A special thanks also goes to my dear friend, Mary Davis, who helped with data collection, editing the manuscript and most of all, gave needed encouragement and support. My warmest appreciation goes to my friend, Michael Eckard, for inspiring me to finish.

The time and expertise of many other individuals was invaluable. An incredible thanks to Dennis Culver and Dave Kerr who so willingly gave their time and energy to help with computer analyses of the data. Thanks also to David Smart for helping with data collection and for the use of the videotapes from his own research. I would also like to express my gratitude for the hard work of Deb Stanislaus for typing and Bonita Kesting for helping prepare tables and figures. Larry Thompson deserves a special word of appreciation for his talent, patience, and assistance. I must also thank all the folks at Boys Town, Achievement Place and Bringing It All Back Home for helping with the data collection phase. In particular, the efforts of Curt Braukmann, Jerry Davis, Gary Timbers, Denny Maloney, Dave Thomas, Dick Baron, Mark Kirchhoff, Tom Dowd and Dan Daly are appreciated. Of course, the research would not have been possible without all those wonderful Teaching-Parents and consultants who donated their time to

participate in the study. Thanks to each and every one of you.

DJS

TABLE OF CONTENTS

List of Tables

LIST OF FIGURES

ABSTRACT

The expansion in social validation literature has initiated a growing concern surrounding several methodology issues. In particular the need has arisen for social validation research to incorporate reliable and valid measurement scales and to explore the effects of extraneous variables on judges' ratings. The current study socially validated the effects of a consultant training program and investigated the effects of four variables on judges' ratings: judges' position, judges' program affiliation, viewed consultant-trainee's level of training and an individual consultant-trainee factor. Teaching-Parent and consultant judges from three different training sites viewed videotapes of consultant-trainees before and after training. The judges rated each trainee's performance using the CPRS, a reliable and valid measurement scale from the counselor literature. Correlational analysis revealed a significant positive relationship between the subjective ratings given to the consultant-trainees and their objective scores (percent appropriate consultant behaviors), thus providing evidence of the social validity of the consultant training program. An analysis of variance showed significant main effects for judges' program affiliation (i.e., training site) and the individual consultant-trainee factor. The variables of judges' position (i.e., Teaching-Parent or Consultant) and consultant-trainee's level of training (i.e., pre or post) did not significantly influence the judges' ratings. Four of the eleven interaction effects were found to be significant. The judges also rated an "ideal" consultant by rating and ranking the CPRS categories. These results suggested the validity and reliability of the CPRS with a consul-

tant population. Results were discussed regarding implications for future social validation research, stressing continued emphasis on social validation methodology.

INTRODUCTION

Statement of Problem and Purpose

Applied behavior analysis research traditionally has focused on determining intervention effects strictly through observational measures and objective data. Using subjective measurement to indicate an intervention's success has been rejected by behavior analysts as being potentially biased, unreliable and invalid. Recently, however, an upsurge of social validation studies has reversed this trend. (Faucett & Miller, 1975; Kent & O'Leary, 1976; Maloney & Hopkins, 1973; Minkin, Braukmann, Minkin, Timbers, Timbers, Fixsen, Phillips & Wolf, 1976; and Willner, Braukmann, Kirigin, Fixsen, Phillips & Wolf, 1977.) Social validation studies involve assessing society's opinions concerning the social acceptability of an intervention program's goals, procedures and/or effects. These studies use subjective reactions of relevant judges (consumers) as valid and, in fact, critical measures of a program's success.

The development of social validation studies is significant for several reasons. First of all, the logic of subjective evaluation in applied research is simple: if applied behavior analysts are creating social programs for society's use, then input from relevant consumers must be tapped or the program may be doomed to fail. Assessing consumer reaction to treatment can provide an important gauge of public support and acceptance of the intervention program. Even if a treatment or a training intervention demonstrates objective behavioral change, the program may not be accepted unless consumers subjectively perceive the recipient of the change as "better". Similarly if the participants themselves are not satisfied with the program, they may voice complaints or actively avoid the

program. The end result is if the program is not accepted, society probably will not utilize the technology, no matter how effective it is, and consequently much time, effort and money is lost (Wolf, 1978).

Secondly, socially validating a treatment program can result in creating a more effective program. For example, there are some data to indicate that a program's effectiveness is increased if the program is acceptable and satisfactory to its participants. Researchers at Achievement Place found that youths (participants) who rated their group home treatment program favorably in the areas of fairness and concern also committed fewer legal offenses during treatment (Braukmann, Kirigin, & Wolf, 1976). This finding suggests that by periodically assessing participants' opinions of a treatment program, program developers can shape the program to adapt more closely to the client's needs. In turn, this may help the clients make important changes that will contribute to their success.

Another stimulus for the increased number of social validation studies is the recent increased sensitivity toward clients' rights (Martin, 1975). As the primary consumers of a treatment program, clients have first-hand knowledge of intervention benefits and/or disadvantages. Since clients have a legal right to the least restrictive alternative in treatment, their opinions should be solicited for ethical reasons. Assessing the opinions of individuals who interact with the client can be another valuable source of socially validating the effects of treatment. As mentioned earlier, even if a client's behavior changes, the changes may not be considered significant by important individuals in the client's life. It has been demonstrated, for example, that if the client is seen as deviant, simply exhibiting nondeviant behaviors may not be sufficient to overcome this

perception (Farina, Allen, & Saul, 1968; Farina, Gliha, Baudreau, Allen & Sherman, 1971). Given these findings and the fact that most interventions aim to improve the client's acceptance by his/her peers and enhance success in interacting with his/her environment, then objective behavioral change may not translate into treatment success. Thus, although treatment may appear successful because of overt, behavioral changes, if significant individuals in the client's life do not subjectively perceive these changes as helpful, treatment success may be limited.

Assessing social validity, therefore, may increase the responsiveness of an intervention program to the needs of its consumers. By requesting consumer opinions an important feedback loop is created that allows program developers to make adaptations which can increase the program's efficiency, effectiveness, usefulness and usability.

Social validation of a treatment or training program can occur on at least three levels: determining the social significance of 1) the goals of the intervention program, 2) the procedures used in the program, and 3) the actual effects or results of the program (Wolf, 1978). Assessing the social validity of the program's goals determines if the targeted behaviors and behavioral goals of the program are considered important ones by significant program consumers. Behavioral goals that have been socially validated through recent research include interaction behaviors of Teaching-Parents preferred by youths in Teaching-Family group homes (Willner, et al., 1977), relevant conversation skills of adolescent girls (Minkin, et al., 1976), appropriate social behaviors of youths interacting with police officers (Werner, Minkin, Minkin, Fixsen, Phillips and Wolf, 1975), the importance of behavioral observation and description skills in

Teaching-Family treatment-related activities (Dancer, Braukmann, Schumaker, Kirigin, Willner, & Wolf, 1978), and the importance of teaching profoundly retarded persons independent walking skills (Gruber, Ruser, & Reid, 1979).

Socially validating the appropriateness of intervention procedures is important in refining the use of behavioral techniques. Although some treatment procedures may be extremely effective, they may be considered unethical, unacceptable, impractical or costly by the clients, caregivers or other consumers. Fox and Azrin (1972), for example, found restitution procedures more acceptable to caregivers than timeout or shock punishment, and reported the use of over-correction procedures with the retarded as acceptable to caregivers. Kazdin (1980) investigated the acceptability of several alternative treatments for deviant child behavior and found reinforcement of incompatible behavior to be the most acceptable treatment followed, in order, by timeout from reinforcement, drug therapy and electric shock. Recently variations of timeout procedures (e.g., "contingent observation", "timeout ribbon") that are less restrictive than traditional timeout rooms have been socially validated as effective and acceptable to users and consumers of various treatment programs (Porterfield, Herbert-Jackson and Risley, 1976; Foxx & Shapiro, 1978).

Finally measuring the social significance of a treatment program's effects allows applied behavior analysts to evaluate consumer satisfaction with both the intended defined results and other unintended results. Validating intervention effects --- that is, determining if consumers feel the treatment program is effective --- has been accomplished through two major study methods: 1) social comparison, and 2) subjective

evaluation (Kazdin, 1977). Social comparison studies compare the client's before and after treatment with the behavior of "nondeviant" peers to determine if the client's behavior after treatment is distinguishable from his/her peers. Social comparison studies using normative data have validated the clinical importance of behavioral change following interventions designed to correct disruptive behavior of children in the classroom and at home (Patterson, 1974; Kent & O'Leary, 1976; Walker & Hops, 1976), to increase social interaction behaviors of isolate children in nursery school (O'Connor, 1972), to improve eating behaviors of adult retardates (Azrin & Armstrong, 1973), to teach conversational behaviors to predelinquent girls (Minkin, et al., 1976), and to teach assertive behaviors to college students (McFall & Twentyman, 1973).

In subjective evaluation studies the client's behavior is evaluated by relevant judges (e.g., individuals likely to have contact with the client, professionals in the field, etc.) to determine the importance of the behavior change made after treatment. These studies assess if the client is viewed by others as qualitatively different or "better" after treatment. Most of these studies entail two steps. First, researchers empirically demonstrate the effectiveness of an intervention through objective behavioral data. Second, the behavioral data are validated by assessing the opinions of relevant consumers through a systematic subjective measure of consumer satisfaction before and after treatment. This second step often involves viewing videotapes of the client's behavior pre-treatment and post-treatment, and then rating the client on a rating scale of one or more dimensions. If improvements in objective data correlate highly with improvements in the subjective ratings then the

effects of the intervention program usually are considered to be socially validated.

Recently, most social validation studies have used the subjective evaluation method to validate treatment or training programs for a variety of populations. A training program for improving conversational skills of predelinquent girls, for example, was validated when judges rated post-training videotapes as reflecting more appropriate conversation than pre-training sessions (Maloney, Harper, Braukmann, Fixsen, Phillips, & Wolf, 1976). In another study with children, Maloney and Hopkins (1973) found that when they modified the sentence structure of stories written by elementary school children, judges' ratings of creativity also increased. In a study with adults, Fawcett and Miller (1975) demonstrated that an instructional package designed to enhance public-speaking behavior effectively increased both objectively-measured public-speaking behavior as well as an audience's subjective ratings of the performance of the trainees. Briscoe, Hoffman & Bailey (1975) validated their training program for teaching adult members of a community board effective problem-solving behaviors when independent judges (professional and community leaders) rated intervention videotapes as portraying greater problem-solving ability than baseline tapes. A social skills training program for modifying interpersonal deficits among retarded adults also was recently validated when relevant judges rated overall interpersonal effectiveness higher in treatment and follow-up videotapes than in baseline videotapes (Bornstein, Bach, McFall, Friman & Lyons, 1980).

With the growing acceptance of social validation as an indicator of a program's success, attention has been turned toward the methodology

of social validation studies. The use of subjective measurement in the applied behavior analysis field currently lacks the rigor and reliability of objective measurement. Kazdin (1977) identified a number of problems with social validation methodology that clearly need to be addressed in future literature. Subjective evaluation studies, for example, usually assess judges' opinions using simple scales designed by the authors. Typically the reliability or validity features of these scales are not tested. Interobserver agreement is the only type of reliability that is assessed and this is done only rarely (Minkin et al., 1976). Face validity is the major criterion for choosing the dimensions that are rated. As studies in psychometric assessment have found, face validity does not assure true validity -- that is, that the scale actually measures what it is designed to measure. If social validity studies are to be continued and useful there is a need for developing and using rating scales which are "tried and true" -- that is, used successfully in several studies. Ideally these scales should be designed and tested on various psychometric dimensions such as test-retest reliability, convergent and discriminant validity, the possibility of set responding, observer bias, etc. Without this advancement in social validation methodology the validity of the results drawn from subjective evaluation studies will continue to be limited by the nature of the measurement system employed.

Another factor that Kazdin identified that may influence the results of subjective evaluation studies is the choice of relevant judges. Most studies employ judges who are not familiar with the clients (Maloney, et al., 1976; Briscoe, et al., 1975; Fawcett & Miller, 1975; Minkin, et al.,

1976; Werner, et al., 1975), although a few have used judges who have met or known the clients personally (Kent & O'Leary, 1976; Dancer, et al., 1978). Choosing judges who are unfamiliar with the client may assure that ratings are not biased by previous perceptions; however, these viewpoints may not be representative of the individuals who have had contact with the client. It could be argued that the perceptions of familiar individuals (e.g., parents, teacher, peers, etc.) are more valid indicators of an intervention program's success because the client's interactions with these individuals after treatment may determine the ultimate success of the program. Since judgement ratings may differ depending upon the choice of judges there is a need for social validation research to address this potentially confounding factor.

Background or environmental variables that may influence judges' ratings have been identified in the counselor literature, which contains a number of studies addressing these questions. Many counselor studies have been designed quite similarly to social validation research: videotapes of counselors or counselor trainees chosen for some objective performance measure or personal characteristic (e.g., amount of training, amount of appropriate counselor behaviors, sex, degree, experience, etc.) are subjectively rated by judges. Counselor studies, however, usually go a step further by varying the judges' experience, background or characteristics (e.g., sex, education, race, job position, experience, etc.) to determine any differential effects on ratings. Occassionally the environmental conditions or standardized instructions will be varied to assess varying effects on ratings (e.g., taped counselor described as experienced versus no experience, Ph.D. degree versus graduate student, etc.) A

brief review of some of the findings of this research may help direct the methodology of future social validation research.

Several variables relating to the judges' (rater's) background or characteristics have been explored as potentially affecting the judges' ratings. Two variables that, in general, have not produced significant differences in judges' ratings include sex (Brown & Cannaday, 1969; Dell & Schmidt, 1976) and race (Peoples & Dell, 1975). The primary variable that has been shown to influence significantly the judges' ratings is the judges' position. Position refers to the judge's job title or relationship to the individual being rated (e.g., counselor, counselor educator, graduate student, supervisor, peer, self, or observer). Conflicting results have been found for this variable. Bishop (1971), for example, compared ratings of counselor effectiveness using counselors, clients and supervisors as judges. Results demonstrated a significant correlation between counselor self-ratings and supervisor ratings. Clients' ratings, on the other hand, did not correlate significantly with either counselor or supervisor ratings, but clients' ratings were significantly higher than the other two groups' ratings. In a similar study, Brown & Cannaday (1969) found different results: a significant correlation was found between counselee and supervisor rankings, whereas counselor self-rankings did not correlate with either of the other two groups. LaCrosse (1977) found significant differences in ratings by clients, observers and counselor self-ratings, with clients giving the highest ratings and observers giving the lowest ratings. Trotzer (1976) found no significant differences between ratings by counselors, counselor educators and graduate

students, but found the ratings of graduate students affiliated with different university programs from different geographical locations to differ significantly. In summary, then, although specific results vary, and in some instances conflict, one might conclude that a judge's position (job title or relationship to individual being rated) may affect the judge's ratings. This finding has implications for the choice of judges in future social validation studies.

Counselor traits (person being viewed or rated) that have afforded significantly different ratings include counselor behavior and an individual counselor factor. Concerning counselor behavior, counselors who use high levels of expert nonverbal behavior (e.g., eye contact, body orientation, etc.) have received significantly higher subjective ratings (Siegel & Sell, 1978; Haase & Tepper, 1972). Also Scheid (1976) found that higher levels of usage of the facilitative core conditions (e.g., counselor warmth, competence, comfort, general appeal, counseling climate and client satisfaction) yield higher subjective ratings. Significant main effects have been found in several studies for an individual counselor factor (Dell & Schmidt, 1976; Peoples & Dell, 1975; Trotzer, 1976). This finding may reflect the existance of some uniformly perceived components of counselor behavior that have not been clearly identified through research to date. The significant findings on these two variables, counselor behavior and individual counselor, have implications for the design of future social validation studies. These results emphasize the importance of clearly identifying behavioral differences in individuals being rated, and analyzing the data which keep individuals separate, rather than grouping which could mask or confound effects. In

general, inconsistent or non-significant findings have been found relating to the effects on ratings due to counselor's sex (Dell & Schmidt, 1976; Brown & Cannaday, 1969; Heppner & Pew, 1977); education level (Engelkes & Roberts, 1970); race (Peoples & Dell, 1975); and level of training or experience (Dell & Schmidt, 1976; Schmidt and Strong, 1970). Replicating these results in future social validation studies with populations other than counselors will be necessary.

The environmental conditions under which the ratings are made also have been manipulated in several counselor studies. Consistent results have been found supporting the hypothesis that higher ratings are given to counselors who are perceived as having higher status. Higher ratings, for example, have been given to counselors who have objective evidence of expertness in the environment, such as diplomas and certificates on their walls (Siegel & Sell, 1978; Heppner & Pew, 1977). This same effect on ratings can be found if the viewed counselors are described in the pre-rating introduction session as having more professional experience, advanced educational degrees and successful reputations (Schied, 1976; Hartley, 1969; Claiborn & Schmidt, 1977). These results have implications for the amount and type of background information that is given to judges concerning the individual being rated in future social validation studies.

In summary the recent upsurge in social validation studies has been influential in increasing the effectiveness, responsiveness and overall acceptance of behavioral intervention programs. A review of the applied behavior analysis literature revealed a number of studies validating the goals, procedures and effects of a variety of treatment or training pro-

grams with several populations. With the expansion in social validation literature comes a growing concern surrounding certain methodological issues. Of particular interest is the need for future social validity research to incorporate reliable and valid scales for measuring subjective opinions. An equally important goal of future research should be to explore the effect of certain extraneous variables on judges' ratings. As revealed in a brief review of the counselor literature, judges' ratings have been shown to differ as a result of the judges' position, the viewed counselor's behavior, the viewed counselor's individuality, and the type of background information given to the judges before rating the individual.

The major purposes of the current research are: 1) to investigate the social validity of the consultation skills taught in a consultant training program using a reliable and valid measurement instrument, and 2) to assess the effects of several variables on the subjective ratings of different groups of relevant judges. Regarding the first purpose, the work is an extension of an earlier study which demonstrated that a training program for consultants to residential child care workers produced changes in the consultant's skill levels (Smart, 1980). The present study assesses the social validity and importance of the consultants' behavior changes as judged by various groups of relevant consumers. It is hypothesized that subjective ratings by relevant judges of consultant-trainees who participated in the consultant training program will correlate positively with the objective measures of these consultant-trainees' skills.

The second purpose addresses some of the methodological issues that

have been problematic in previous social validation research. The current research particularly addresses the concern that social validation results may vary depending upon the audience whose opinions are solicited and depending upon the characteristics or behavior of the individual viewed. The specific variables investigated include 1) judge's job position (Teaching-Parent or Consultant), 2) judge's program affiliation (Site #1, #2 or #3), 3) the level of viewed consultant's training (pre- or post-training), and 4) an individual viewed consultant factor (Trainee #1, #2, #3, or #4). It is hypothesized that the subjective ratings of the viewed consultant-trainees will differ as a result of these four variables.

METHOD

Setting

The Teaching-Family Model is a child-care treatment model used in community-based and campus-based group homes for children and adolescents with problems. The Teaching-Family Model is a professional, positive and practical model of child care that teaches troubled children alternative, more appropriate behaviors that will aid their successful reintegration into family and community life. Several key components of the Teaching-Family treatment program include teaching, family-style living, relationship development, a token economy motivation system, a self-government system, and a strong family and community orientation.

There are approximately 150 Teaching-Family group homes located across the nation, and most of these are affiliated with one of several major Teaching-Family sites. A Teaching-Family site is a regional training center that sponsors several group homes in one general geographical location. The sponsor site is responsible for providing training, consultation and evaluation services for the residential child-care workers. The child-care workers in Teaching-Family group homes are called Teaching-Parents. The Teaching-Parents are a married couple that live in the home with the youths and are directly responsible for implementing the Teaching-Family treatment program with each individual youth. Consultants are members of the training site staff who provide direct feedback and advice to the Teaching-Parents concerning their implementation of the Teaching-Family program.

The current study was conducted at three Teaching-Family Regional

Training Centers (Sites) across the country: Father Flanagan's Boys' Home at Boys Town, Nebraska; Achievement Place at Lawrence, Kansas; and Bringing It All Back Home at Morganton, North Carolina.

Background

Since the current research is an extension of an earlier study by Smart (1980), a brief description of this earlier study is provided below. The cited study described the development of a training program for consultants and demonstrated the effectiveness of the training program in teaching consultants specific procedures and techniques. The consultant-trainees were asked to role-play a number of behavior simulations before and after training. These simulations were designed to portray situations the consultant would likely encounter during actual consultation sessions with Teaching-Parents. The simulations were videotaped, and then the pre- and post-videotapes were scored using checklists of previously determined appropriate and inappropriate consultant behavior. Results of the study showed that consultants used a higher percentage of appropriate consultant behaviors and a lower percentage of inappropriate behaviors in the post-training tapes, thus providing evidence of the effectiveness of the training program. The current research extends the Smart (1980) study by investigating the social validity of the skills taught in the consultant training program.

Judges

Forty-six staff from three different Teaching-Family sites (Regional Training Centers) across the country volunteered to serve as judges. Each

site uses a consultation system similar to the one described by Smart (1980). There were 14 judges from Site #1, 16 from Site #2, and 16 from Site #3. Judges were categorized into two groups according to their job position: Teaching-Parent or consultant. From Site #1 there were 8 Teaching-Parents and 6 consultants, from Site #2 there were 8 Teaching-Parents and 8 consultants, and from Site #3 there were 8 Teaching-Parents and 8 consultants.

Teaching-Parents and consultants were selected to provide a relevant sample of judges. Teaching-Parents were selected who had less than one year's experience and who received consultation services at the time of the study. These Teaching-Parents were viewed as the most relevant consumers of the consultant training program since Teaching-Parents usually receive consultation services more frequently during their first year. The average number of months experience for the Teaching-Parent judges from Site #1 was 7.25 months, from Site #2 was 10.5 months, and from Site #3 was 6.5 months. Experienced consultants were selected as representing the most relevant users of the consultant training program. Only consultants who were currently consulting or had consulted in the recent past were asked to participate. The mean number of months of consulting experience for the judges from Site #1 was 40.0 months, from Site #2 was 29.8 months, and from Site #3 was 10.1 months. It should be noted that none of the consultant judges had participated in the new consultant training program described by Smart (1980) since this experience and knowledge was considered a possible confound.

The judges ranged in age from 21 to 44 years, with a mean age of 27.0 years for Teaching-Parents and 30.6 years for consultants. There were

23 females (12 Teaching-Parents and 11 consultants) and 23 males (12 Teaching-Parents and 11 consultants) in the study. Eleven consultant judges held B.A. or B.S. degrees, five held M.A. degrees, and one held a Ph.D. degree. Of the Teaching-Parent judges, 1 had a high school education, 5 attended college but did not receive a degree, 11 held B.A. or B.S. degrees, and 3 held M.A. degrees. (Educational information was not retrievable for nine judges).

Procedures

General Procedures. In the present study pre- and post-training videotapes from a random sample of the consultant-trainees from the Smart study (1980) were shown to the judges. The judges were asked to rate the videotaped consultant-trainees using a reliable measurement instrument from the counselor literature.

In the experimenter's initial contacts with the judges, the judges were asked to complete a background information sheet which included information such as age, sex, educational background, experience in the Teaching-Family Model and consulting experience. Each judge and each videotaped consultant trainee also completed a written informed consent. Samples of the Participant Background Information Form and Informed Consent Forms can be found in Appendix A.

The data were collected between November, 1978 and January, 1979 at each of the three sites. Videotapes were shown separately to groups of Teaching-Parents and consultants at each site. Each session took approximately two hours and judges received $5.00 for their participation.

Before viewing the videotapes, the judges were given a general de-

scription of the study and specific standardized verbal instructions concerning how to use the measurement scale. (See Appendix B for Experimenter's standardized instructions.) Standardized instructions were used to minimize differences in environmental conditions. This was considered necessary since the data were collected at different sites and times. In addition, previous counselor studies have found ratings to be affected by the amount and type of background information given about the viewed counselors (Hartley, 1969; Schied, 1976; Claiborn & Schmidt, 1977). In this study judges received no background information on the videotaped consultant-trainees. Also, judges did not know whether they were viewing a pre-training or a post-training videotape.

Designing the Videotape Sample

Description of Videotaped Simulations.

In the Smart (1980) study consultant-trainees role-played four different behavioral simulations designed to measure three different consultant behavior areas. Simulation #1 measured the consultant's ability to initiate relationships with their clients, Simulations #2 and #3 assessed the consultant's observation-visit skills, and Simulation #4 (Part A and B) assessed the consultant's ability to give feedback and advice to Teaching-Parents. In the present study a limited number of these simulations were shown to the judges since a lengthy viewing session could result in fatigue which could influence or invalidate judges' ratings.

Simulation #4 (Part A) was chosen for viewing because 60% of the training time was spent teaching how to give feedback and advice. This simulation, therefore, best represented the skills taught in the consulta-

tion workshop. In this simulation a Teaching-Parent actor initiates discussing a problem with the consultant-trainee concerning a youth who is described as either shy and unresponsive or stubborn and moody. The consultant's task is to ascertain a behavioral description of the child's problem from the Teaching-Parent and to help design a plausible problem-solving plan with the Teaching-Parent.

Sample Selection: Choice of Consultant-trainee Videotapes. In the Smart (1980) study twelve consultant trainees were videotaped in Simulation #4 (Part A) before and after training, resulting in a total of 24 videotapes. For the current study, eight videotapes were randomly chosen from the twenty-four. It was estimated that it would take approximately two hours for each judge to view the eight videotapes -- a time allottment which would not produce a debilitating amount of fatigue.

To ensure that a representative sample of consultant skill levels would be presented, one restriction was placed on the random sampling procedure. Consultant-trainee's skill levels were determined in the Smart (1980) study by objectively measuring the consultant's appropriate and inappropriate behaviors. Specifically, each pre- and post-videotape was scored using the checklist shown in Appendix C. The sampling restriction involved choosing one quarter of the sample, or two videotapes, from pre-training videotapes in which the consultant-trainees' objective scores were below the overall median score. Another quarter of the sample was chosen from consultant-trainees with pre-scores above the median score for the entire group. After these four pre-training videotapes were chosen, their corresponding post-tapes were selected thus assuring that the judges would view the same consultant-trainee's pre- and post-train-

ing videotape. Pairing the pre- and post-tapes was done to allow analysis of any effects due to an individual consultant-trainee factor. This analysis was deemed important since several counselor studies showed an individual counselor factor contributed to significant differences in ratings (Peoples & Dell, 1975; Dell & Schmidt, 1976; Trotzer, 1976). The objective performance scores of the four consultant-trainees chosen are exhibited in Table I in Appendix D. To determine if the consultant skill scores of the subsample were statistically comparable to the total sample a t-test was performed. Results demonstrated the groups were not significantly different, providing some support for considering the groups comparable. Results of the t-test as well as a comparison of sex of the subsample with the total sample, are located in Tables II and III in Appendix D.

Control for Order Effects. It was considered necessary to control for a possible order effect since there were eight different videotapes to be viewed (one pre- and one post-videotape for each of the four trainees). Accomplishing this entailed assembling and transcribing the eight videotapes onto two different master videotapes. The first master videotape contained the eight scenes sequenced in a random order, designated as regular order. The order was selected randomly with the restriction that the same consultant-trainee not be seen consecutively. Then the numbers one through eight were assigned to each of the eight videotaped segments on the regular-sequenced master videotape. The second master videotape contained the same eight videotaped segments, but in a counterbalanced order, with segment five through eight shown first and segment one through four shown next.

Half of the judges viewed the regular order tape and half viewed the counterbalanced order tape. The order in which a judge viewed the tapes was randomly assigned with the restriction that half of the Teaching-Parents from each site viewed regular order and half viewed counterbalanced order. The same restriction was placed on the consultant judges from each site. A t-test was performed to ascertain if there were significant differences between the ratings of judges viewing regular order and those viewing the counterbalanced order. The results were non-significant, indicating that counterbalancing had achieved its purpose and order of viewing did not significantly affect the ratings. The t-test results are represented in Table IV in Appendix E.

The Measurement Instruments

The Counselor Performance Rating Scale (CPRS), developed originally by Kelz (1966) to measure counselor performance, was the major measurement instrument used in the current study. This scale was chosen for several reasons. First, a literature review did not reveal a measurement scale strictly for consultant behavior, thus the counselor literature seemed to afford the best alternative. The only necessary change made in the CPRS was replacing each instance of the word "counselor" with "consultant".

Second, since the primary focus on the study was on differences in viewers' impressions, a rating scale was needed that would be broad enough to detect these differences. The CPRS is a 5-point scale, ranging from "unsatisfactory" (rating of "1") to "outstanding" (rating of "5"). Another advantage of this ordinal scale was that analysis through para-

metric statistics would be possible.

Third, the nature of the study dictated that a relatively short measurement instrument be used in the interest of saving time. Long, detailed scales can lose judges' interest and consequently invalidate ratings. The CPRS, being an 8-item scale, was considered to be an appropriate length. An equally important reason for choosing the CPRS was the content and wording of the categories on the scale. The face validity of the CPRS for the consultant population was considered as excellent, since the eight categories tapped characteristics and qualities deemed appropriate for good consultants. The eight categories include appearance, expression, relationship, communication, knowledge, perception, interpretation and termination. The CPRS also provided subcategories and defined each category to aid judges' ratings. It was felt this would be helpful in standardizing the judges' frame of reference so that differences in ratings would reflect true differences in perception. (See Appendix F for a copy of the CPRS and the definitions for its use.)

The final reason for choosing the CPRS was that the CPRS has been used successfully in several studies in the counselor literature, and consequently, acceptable reliability and validity data were available for the scale (Patterson, 1966; Kelz, 1966; Johnson & Fredrickson, 1968; Ryan, Johnson, Folsom & Cook, 1970; Thomas & Britton, 1971). These reliability and validity results are summarized in Table V in Appendix G. Using an established scale with available reliability and validity data appears to offer a needed degree of rigor to the social validity methodology.

As mentioned earlier the CPRS was completed by <u>each judge</u> for <u>each of the eight videotaped segments</u> of the consultant trainees. Two addi-

tional measurements were included to facilitate assessing the validity of the CPRS with the consultant population. Before rating the videotapes the judges were asked to think about how an "ideal" consultant would discuss a youth problem with a Teaching-Parent. The judges were then asked to rate this "ideal" consultant using the CPRS. Theoretically, if the CPRS was a perfect evaluation tool for our consultant population, the ideal consultant should be rated a perfect "5" ("outstanding") in all eight categories. It would seem that the closer the ideal consultant was rated to 5.0, the more valid the CPRS is for the current study. The scoresheet used to rate the ideal consultant is contained in Appendix H.

After rating the videotapes, the judges also were asked to rate and rank each of the eight major categories on the CPRS concerning how important each category was in discriminating a good consultant. The categories were ranked with "1" being most important and "8" being least important. Ratings were given on a 7-point scale where "7" indicated that the category was "completely important" for discriminating a good consultant, while "1" indicated the category was "completely unimportant". Results were used to establish the validity of using the CPRS with the current population. (See Appendix I for a copy of the instructions and scoresheet used for rating and ranking the CPRS itself.)

The final measurement taken from each judge was an assessment of the judges' familiarity with the videotaped consultant-trainees. Although this measurement was demographic in nature, it was taken to address Kazdin's (1977) concern that ratings may vary depending on the judges' familiarity with the client or trainee. Familiarity was assessed based on the number of personal social contacts and the number of consulting

contacts a judge had experienced with the consultant-trainee. A copy of the two questions used to determine a judge's familiarity with the consultant-trainees can be found with the CPRS in Appendix F.

Experimental Design and Analysis

The current study used a mixed experimental design with two between-subjects' variables and two within-subjects' variables (Myers, 1967). The two between-subjects' variables included: 1) judges' position (Teaching-Parent or consultant) and 2) judges' program affiliation (Site #1, Site #2 or Site #3). The two within-subjects' variables included: 1) the level of consultant-trainee training (pre-training or post-training), and 2) an individual consultant trainee factor (Trainee #1, Trainee #2, Trainee #3 and Trainee #4). A pictoral representation of the design is located in Figure 1.

- - - - - - - - - - - - - - - -
Insert Figure 1 about here
- - - - - - - - - - - - - - - -

The major statistical analysis performed was an analysis of variance with repeated measures and unequal n's. The analysis of variance tested how the four major independent variables listed above influenced the judges' ratings. The dependent variable was the judges' ratings on the CPRS. Each judge used the CPRS to rate all four consultant-trainees under both training conditions (pre and post) resulting in eight total viewings. For each of the eight videotaped segments viewed, a judge's ratings on all eight categories of the CPRS were totaled and divided by eight. This computation yielded an average overall rating for that particular judge for each individual consultant-trainee. These mean ratings formed

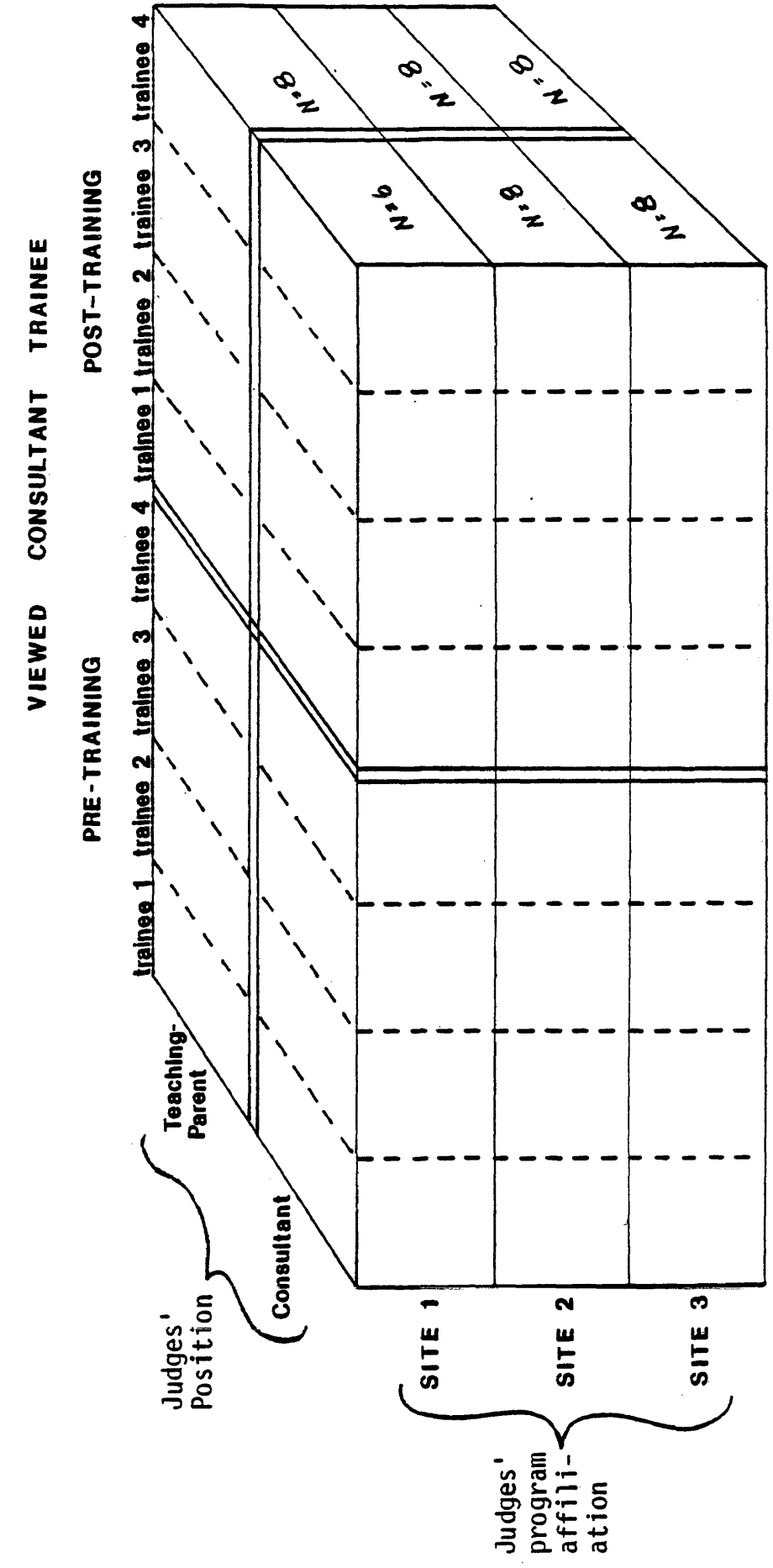

Figure 1. Pictorial representation of Design: Mixed design with 2 between-subjects variables and 2 within-subjects variables.

the basis for the analysis of variance (a summary of these raw data is presented in Table VI in Appendix J).

The second major analysis performed was an analysis of covariance using the judges' ratings of an "ideal" consultant as the covariate. This analysis tested whether or not a judge's ratings of an "ideal" consultant covaried with and, consequently, influenced the outcome of the analysis of variance.

The final statistical analysis performed was a Kendall rank correlation (Wike, 1971). This analysis assessed the possibility of a significant positive correlation between judges' ratings on the CPRS and the objective measurement of a consultant-trainee's performance (e.g., the percentage of appropriate consultant behaviors scored on the checklist in Appendix C from the Smart (1980) study). It was hypothesized that as a consultant-trainee's objective performance improved (higher percentage of appropriate consultant behavior scored), the subjective ratings given by the judges also would increase. (A summary of each consultant-trainee's objective performance scores was previously presented in Table I in Appendix D.) The .05 level was set as the minimum significance level for all statistical tests.

Reliability Checks and Missing Data. The complexity of the design and the vast amount of data necessitated analysis by computer. Assessing the reliability of entering the raw data into the computer involved spot checking the accuracy on 43% of the entire data set. Only nine errors were found (less than 0.5% of the data set), and these were corrected before the analyses were run. Due to judges' scoring errors, there were twelve scores missing out of a total possible 4048 scores. These missing

data constituted only 0.3% of the entire data set and therefore were considered insignificant.

RESULTS

The mean CPRS ratings for all four independent variables (site x position x consultant trainee x training level) are portrayed in Figure 2. The results of the analysis of variance for these four independent variables were presented in Table VII. As predicted, judges' ratings did differ significantly as a result of some, but not all, of these variables. Significant differences were found in the judges' ratings based on the judges' program affiliation. This significant main effect for site ($p < .001$) is depicted graphically in Figure 3. The mean overall rating on the CPRS for Site #1 was 3.58, for Site #2 was 2.97 and for Site #3 was 3.79. A significant main effect was also found for an individual consultant-trainee factor ($p < .001$). Figure 4 displays the mean rating given to each of the four consultant-trainees viewed, with Trainee #1 receiving a mean rating of 3.30; Trainee #2, a mean rating of 2.85; Trainee #3, a 3.88; and Trainee #4, a 3.75. There were no significant differences found due to the judges' position (Teaching-Parent or consultant) or the consultant-trainee's level of training (pre or post).

- - - - - - - - - - - - - - - -
Insert Figure 2 about here
- - - - - - - - - - - - - - - -

- - - - - - - - - - - - - - - -
Insert Table VII about here
- - - - - - - - - - - - - - - -

- - - - - - - - - - - - - - - -
Insert Figure 3 about here
- - - - - - - - - - - - - - - -

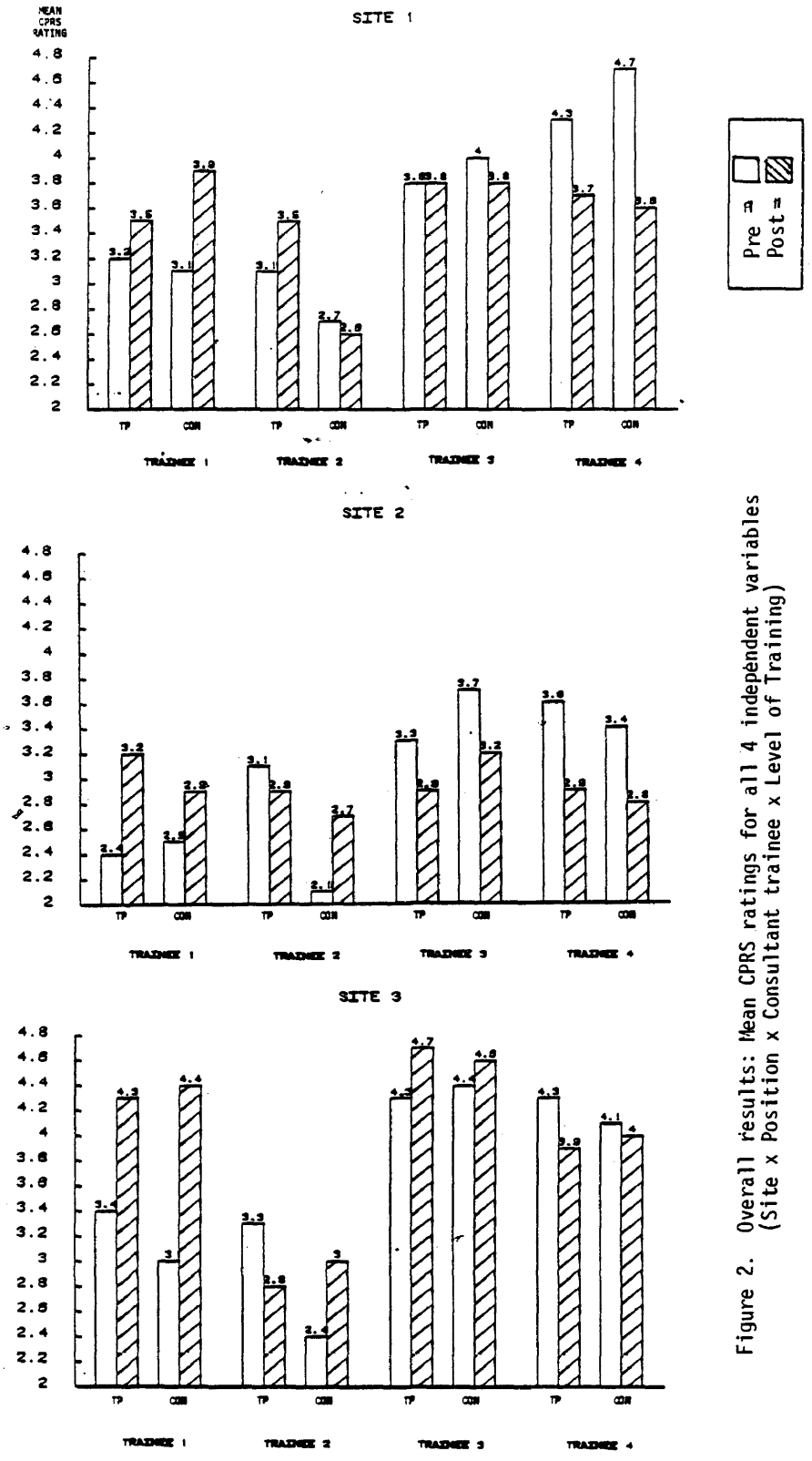

Figure 2. Overall results: Mean CPRS ratings for all 4 independent variables (Site x Position x Consultant trainee x Level of Training)

SOURCE	SUM OR SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROBABILITY
MEAN	4321.29381	1	4321.29281	3992.29	0.0000
POSITION	1.04018	1	1.04018	.96	.3328
SITE	46.38773	2	23.19386	21.43	.0000**
POSITION x SITE	.06492	2	.03246	.03	.9705
ERROR	43.29635	40	1.08241		
TRAINEE	61.48844	3	20.49615	38.79	0.0000**
TRAINEE x POSITION	5.26147	3	1.75382	3.32	.0222*
TRAINEE x SITE	12.12651	6	2.02108	3.82	.0016**
TRAINEE x SITE x POSITION	1.89819	6	.31637	.60	.7309
ERROR	63.40990	120	.52842		
TRAINING LEVEL	.26004	1	.26004	.68	.4159
TRAINING LEVEL x POSITION	.32439	1	.32439	.84	.3641
TRAINING LEVEL x SITE	2.60682	2	1.30341	3.39	.0438*
TRAINING LEVEL x POSITION x SITE	1.19917	2	.59958	1.56	.2231
ERROR	15.39385	40	.38485		
TRAINEE x LEVEL OF TRAINING	21.97143	3	7.32381	22.41	.0000**
TRAINEE x LEVEL OF TRAINING x POSITION	1.29406	3	.43135	1.32	.2711
TRAINEE x LEVEL OF TRAINING x SITE	2.40258	6	.40043	1.23	.2981
TRAINEE x LEVEL OF TRAINING x POSITION x SITE	3.40086	6	.56681	1.73	.1187
ERROR	39.21573		.32680		

** $p < .01$

* $p < .05$

Table VII. Analysis of Variance Results: Differences in Ratings of Consultant Trainees Before and After Training by Teaching-Parent and Consultant Judges from Three Different Sites.

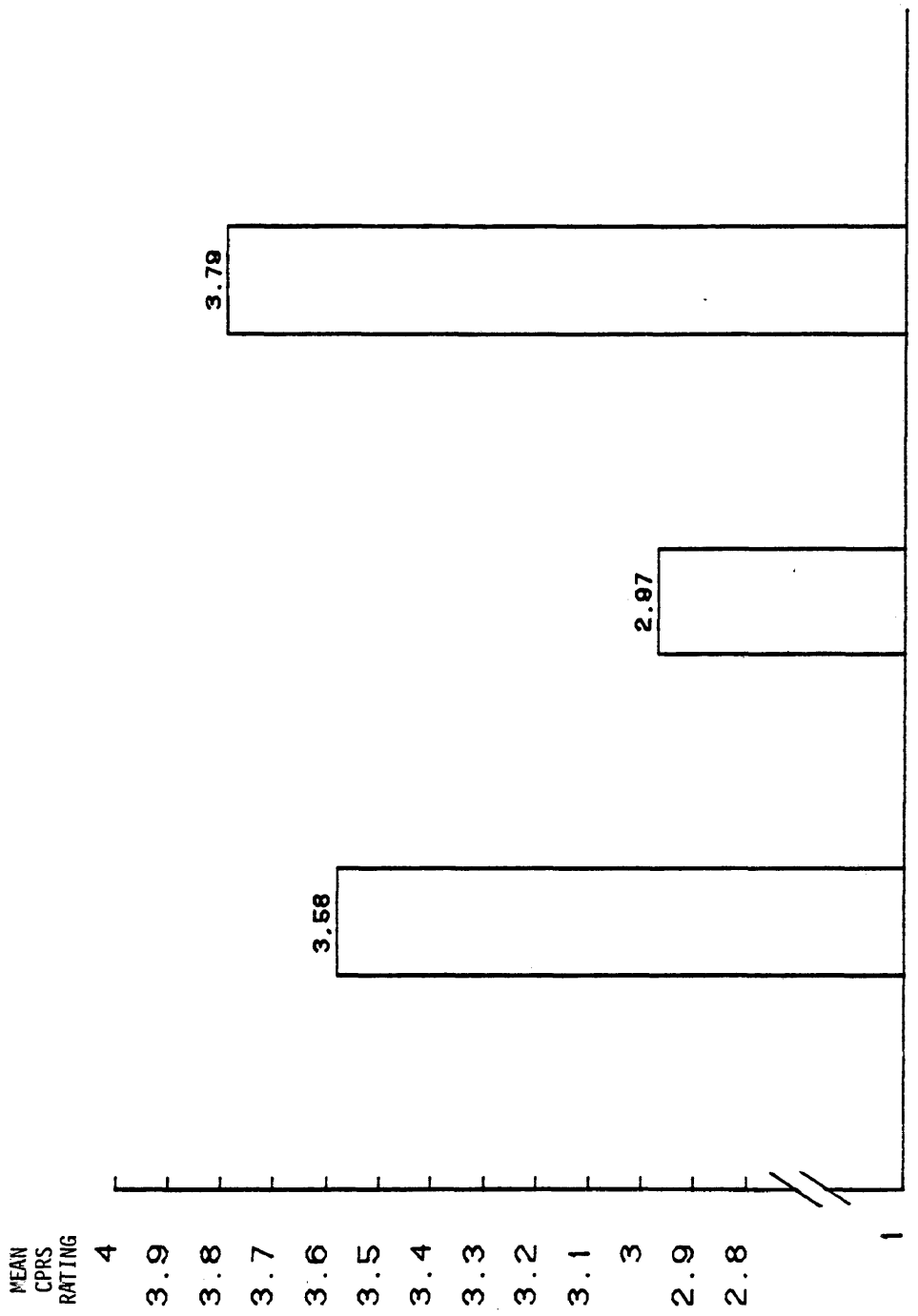

Figure 3. Mean overall rating on CPRS for all video taped consultant-trainees as judged by three regional training sites: a significant main effect ($p>.001$)

- - - - - - - - - - - - - - - -
Insert Figure 4 about here
- - - - - - - - - - - - - - - -

The analysis of variance also revealed several two-way interaction effects between variables. A significant interaction effect was found between the site factor and the individual consultant-trainee factor ($p < .01$). This interaction effect is depicted graphically in Figure 5. Two other interaction effects were significant at the $p < .05$ level: The interaction between judges' position and individual consultant-trainee and the interaction between individual consultant-trainee and the trainee's level of training. Both of these interactions are graphed in Figures 6 and 7, respectively. A final interaction effect was noted between the variables of judges' program affiliation (site) and consultant-trainee's level of training (pre or post). This interaction was also significant at the $p < .05$ level and is represented pictorially in Figure 8.

- - - - - - - - - - - - - - - -
Insert Figure 5 about here
- - - - - - - - - - - - - - - -

- - - - - - - - - - - - - - - -
Insert Figure 6 about here
- - - - - - - - - - - - - - - -

- - - - - - - - - - - - - - - -
Insert Figure 7 about here
- - - - - - - - - - - - - - - -

- - - - - - - - - - - - - - - -
Insert Figure 8 about here
- - - - - - - - - - - - - - - -

The results of the analysis of covariance are represented in Table VIII. Inspection of this table reveals that the covariate (e.g., ratings of "ideal" consultant on CPRS) did not significantly effect the results. In fact, exactly the same main effects and interaction effects were found

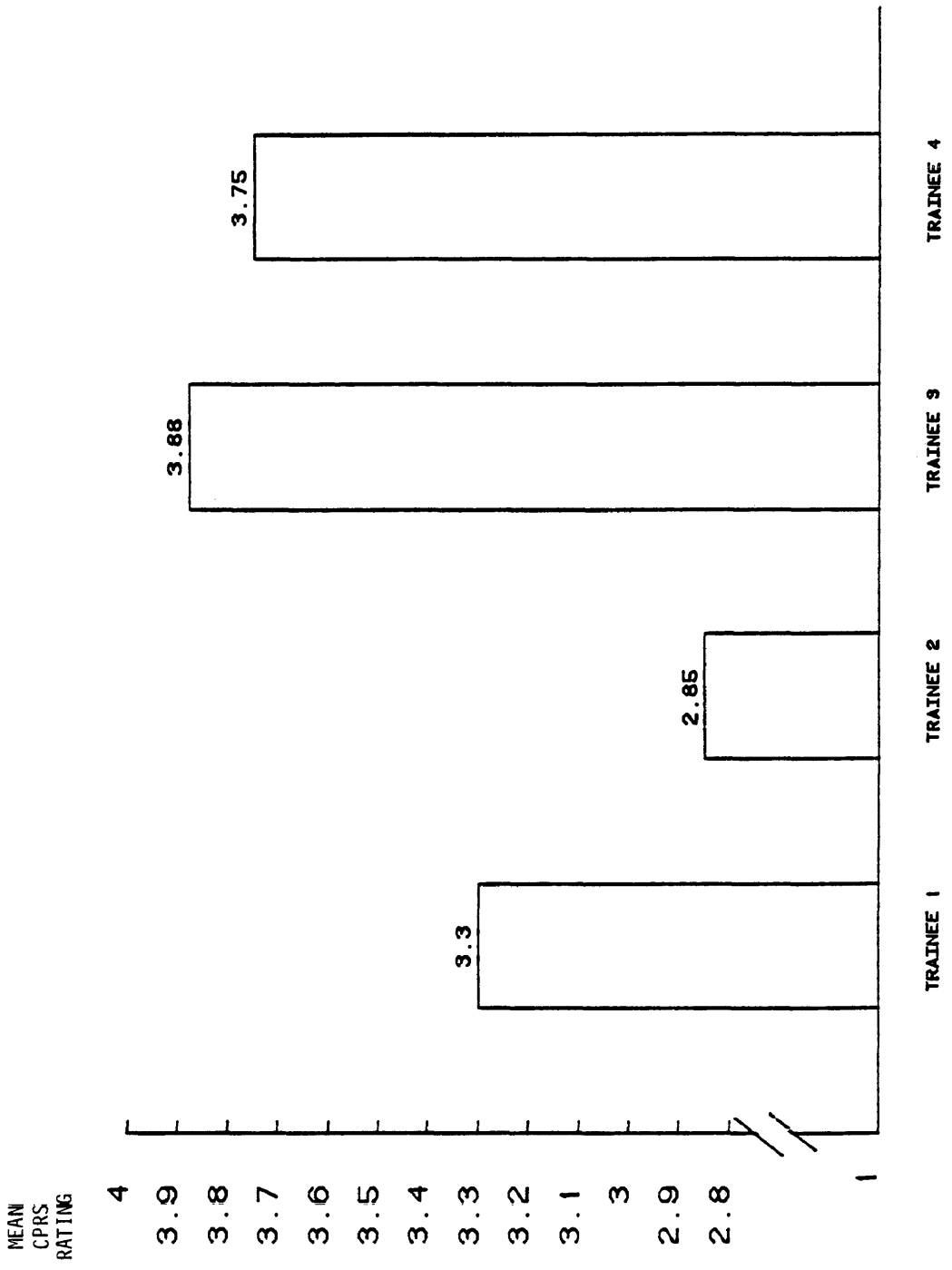

Figure 4. Mean rating on CPRS for each individual video taped consultant-trainee: a significant main effect ($p > .001$)

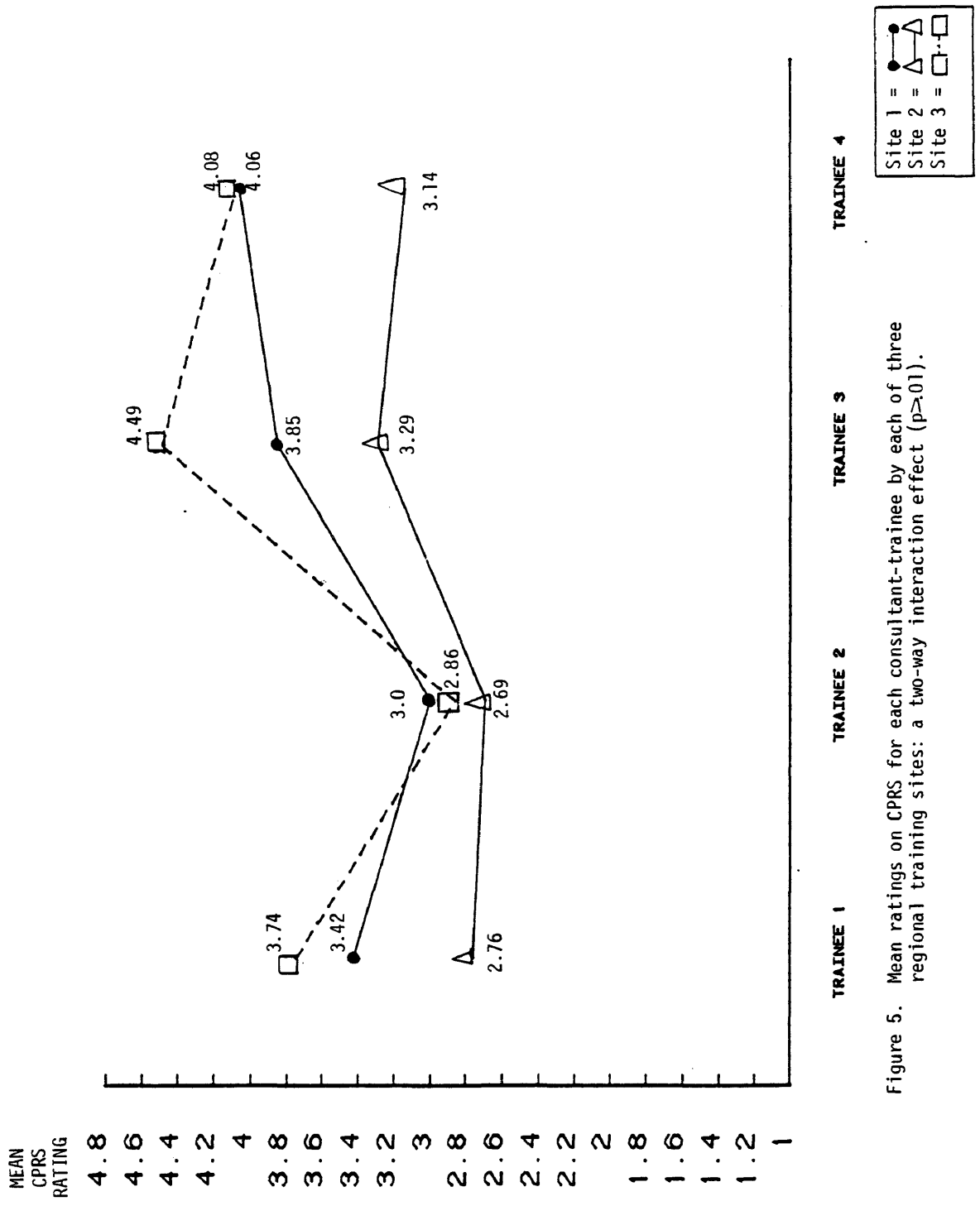

Figure 5. Mean ratings on CPRS for each consultant-trainee by each of three regional training sites: a two-way interaction effect ($p \geq .01$).

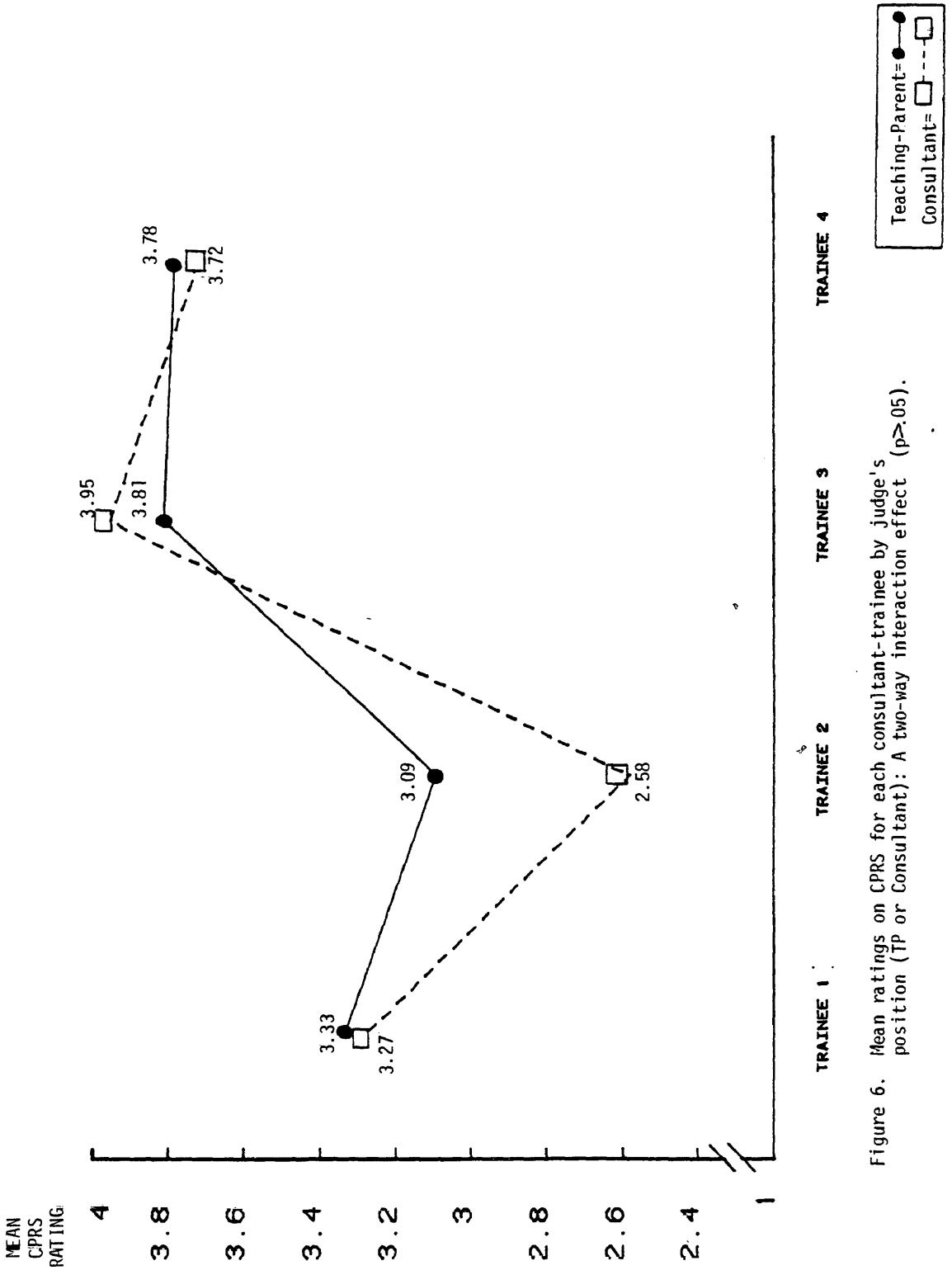

Figure 6. Mean ratings on CPRS for each consultant-trainee by judge's position (TP or Consultant): A two-way interaction effect ($p>.05$).

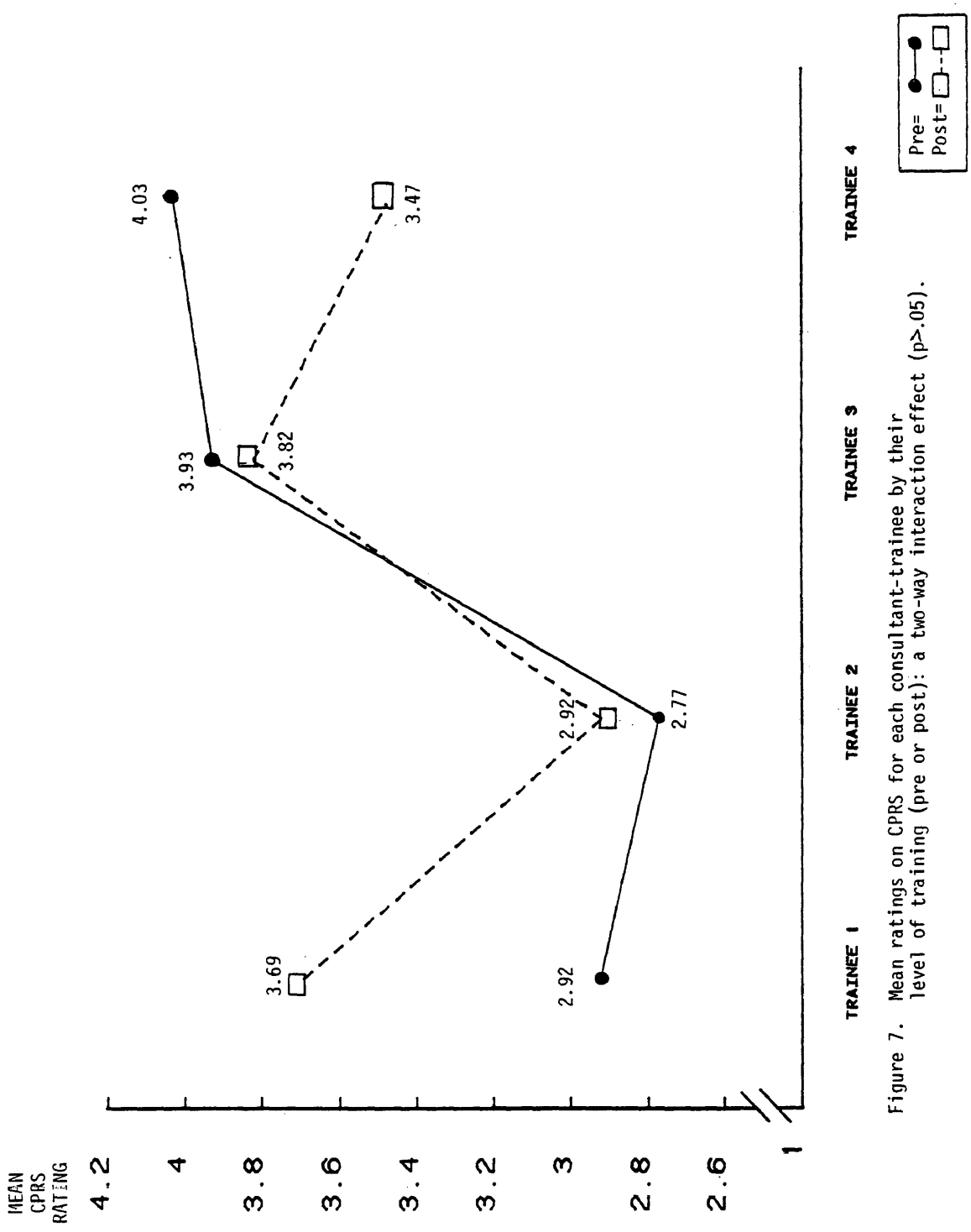

Figure 7. Mean ratings on CPRS for each consultant-trainee by their level of training (pre or post): a two-way interaction effect ($p>.05$).

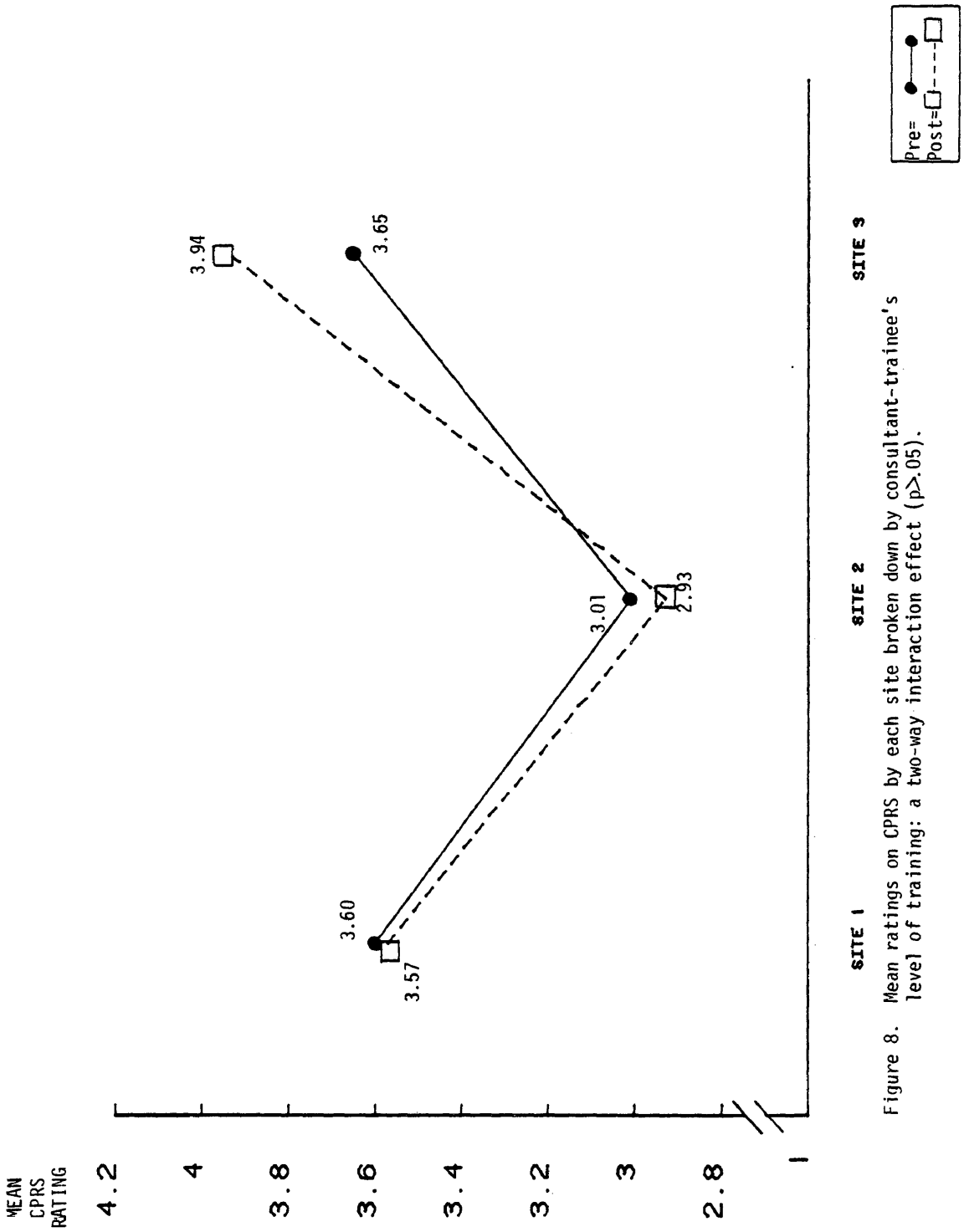

Figure 8. Mean ratings on CPRS by each site broken down by consultant-trainee's level of training: a two-way interaction effect ($p>.05$).

as in the original ANOVA.

- - - - - - - - - - - - - - - -
Insert Table VIII about here
- - - - - - - - - - - - - - - -

The results of the Kendall correlation coefficient supported the hypothesis that consultant-trainees who earned higher objective scores would also receive higher subjective ratings on the CPRS by the judges. Kendall's Tau indicated a small, but significant, positive correlation of .23 ($p < .001$). This positive linear relationship is represented graphically in Figure 9 depicting how the objective scores and subjective ratings increase together.

- - - - - - - - - - - - - - - -
Insert Figure 9 about here
- - - - - - - - - - - - - - - -

The results of the two measures taken to assess the validity of using the CPRS with a consultant population are summarized in Tables IX, X and XI. Table IX reflects the mean CPRS ratings of an "ideal" consultant as viewed by Teaching-Parent and consultant judges from each of the three sites. From Site #1, Teaching-Parents rated an "ideal" consultant at 4.56, while the consultants gave a mean rating of 4.68. From Site #2 both Teaching-Parents and consultants gave a mean rating of 4.55. The Teaching-Parents from Site #3 rated an "ideal" consultant on the average at 4.54 and the consultants at 4.63.

- - - - - - - - - - - - - - - -
Insert Table IX about here
- - - - - - - - - - - - - - - -

The data concerning the rating and ranking of the CPRS scale's ability to discriminate a good consultant are summarized in Tables X and XI. In Table X mean ratings on a 7-point scale (with "1" representing

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROBABILITY	BETA ESTIMATES
MEAN	15.53366	1	15.53366	14.23	.0005	
POSITION	1.21472	1	1.21472	1.11	.2980	
SITE	45.57886	2	22.78943	20.88	.0000	
POSITION x SITE	.07224	2	.03612	.03	.9675	
1-ST COVAR (Ideal)	.72311	1	.72311	.66	.4207	.13312
ERROR	42.57325	39	1.09162			
TRAINEE	61.48844	3	20,49615	38.79	0.0000	
TRAINEE x POSITION	5.26147	3	1.75382	3.32	.0222	
TRAINEE x SITE	12.12651	6	2.02108	3.82	.0016	
TRAINEE x POSITION x SITE	1.89819	6	.31637	.60	.7309	
ERROR	63.40990	120	.52842			
TRAINING LEVEL	.26004	1	.26004	.68	.4159	
TRAINING LEVEL x POSITION	.32436	1	.32439	.84	.3641	
TRAINING LEVEL x SITE	2.60682	2	1.30341	3.39	.0438	
TRAINING LEVEL x POSITION x SITE	1.19917	2	.59958	1.56	.2231	
ERROR	15.39385	40	.38485			
TRAINEE x TRAINING LEVEL	21.97143	3	7.32381	22.41	.0000	
TRAINEE x TRAINING LEVEL x POSITION	1.29406	3	.43135	1.32	.2711	
TRAINEE x TRAINING LEVEL x SITE	2.40258	6	.40043	1.23	.2981	
TRAINEE x TRAINING LEVEL x POSITION x SITE	3.40086	6	.56681	1.73	.1187	
ERROR	39.21573	120	.32680			

Table VIII. Analysis of Covariance Results with Ratings of "Ideal" Consultant as Covariate.

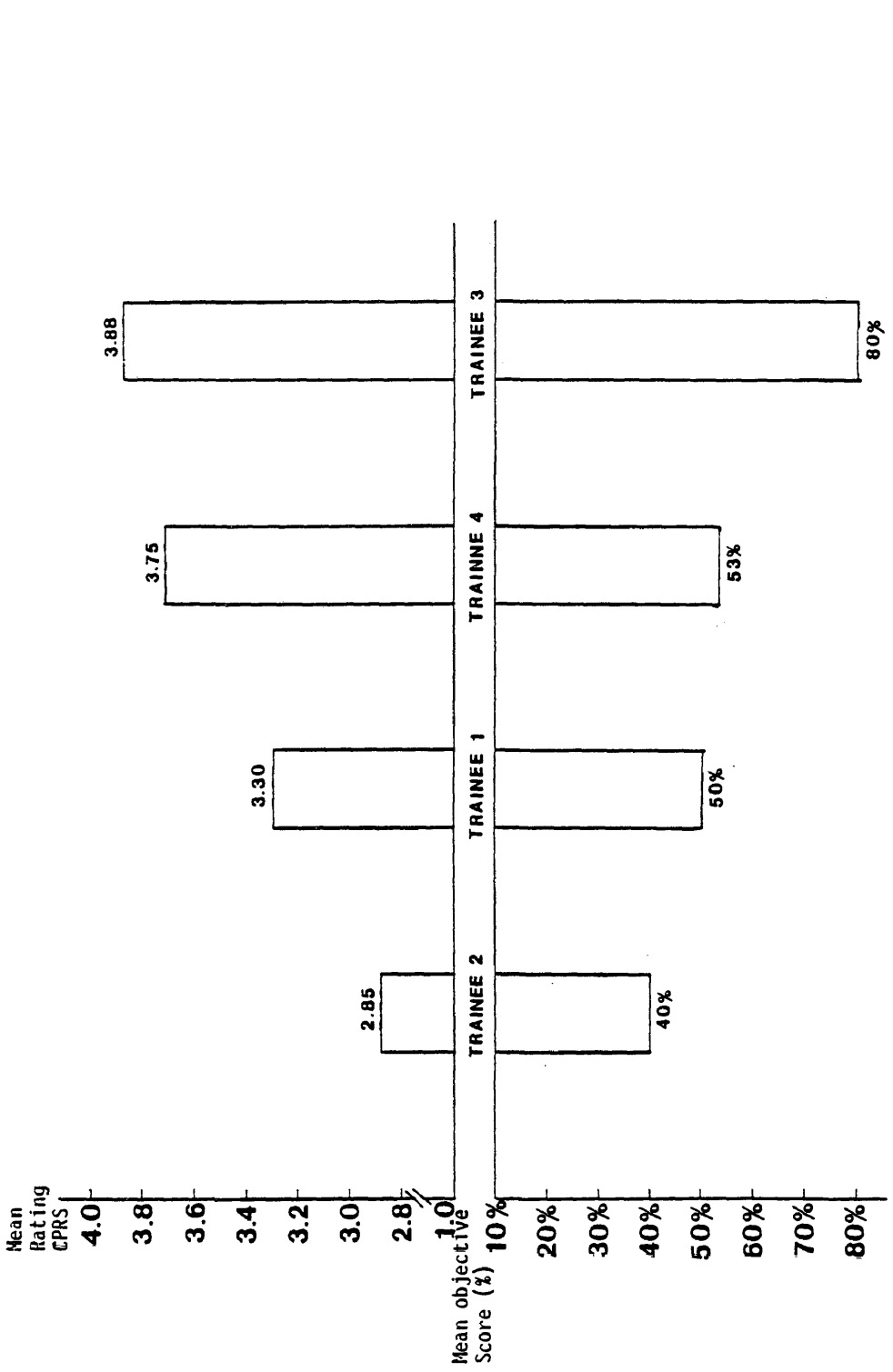

Figure 9. Pictoral representation of Kendall's Teu: Mean ratings on CPRS for each consultant-trainee as correlated with the trainee's percentage of appropriate consultant behaviors (objective score)(p>.001

TABLE IX

Mean Ratings on CPRS Scale for "Ideal" Consultant by Teaching-Parents and Consultants at 3 Regional Teaching-Family Sites

Position	Site 1		Site 2		Site 3	
	$\bar{x}$	S.D.	$\bar{x}$	S.D.	$\bar{x}$	S.D.
Teaching-Parent	4.56	.378	4.55	.278	4.54	.447
Consultant	4.68	.402	4.55	.312	4.63	.311

"completely unimportant" and "7" representing "completely important") are displayed for each of the eight categories on the CPRS scale. Again, the data are organized by judges' position (Teaching-Parent or consultant) and program affiliation (site). The mean ratings ranged from 4.1 for the category Appearance to 6.7 for the category Knowledge, with 5 out of the 8 categories receiving an average rating greater than 6.0 ("Important"). The overall mean rating for the CPRS collapsing all categories was 5.83. Table XI summarizes how the judges from each site ranked each of the CPRS scale categories, with a ranking of "1" being the most important category in discriminating a good consultant. Combining all the judges, the CPRS categories were ranked in the following order of descending importance: Knowledge, Relationship, Perception, Interpretation, Concern, Expression, Termination and Appearance.

- - - - - - - - - - - - - - -
Insert Table X about here
- - - - - - - - - - - - - - -

- - - - - - - - - - - - - - -
Insert Table XI about here
- - - - - - - - - - - - - - -

The results of the final measurement, the assessment of each judges' familiarity with the consultant-trainees, are exhibited in Table XII. Regarding personal contacts, from Site #1 most of the judges were unfamiliar with all four consultant-trainees, with the exception that two had met Trainee #2, two had met Trainee #3, and two had met Trainee #4. None of the judges from Site #2 knew any of the consultant-trainees. From Site #3, on the other hand, 5 judges knew Trainee #1, 9 judges know Trainee #2, and 14 judges know Trainee #3 and Trainee #4. Concerning consulting experience with the consultant-trainees, the judges from Site #1 and Site #2 had no

CPRS CATEGORIES

	APPEARANCE	EXPRESSION	RELATIONSHIP	CONCERN	KNOWLEDGE	PERCEPTION	INTERPRE-TATION	TERMINATION
SITE 1 ($\bar{x}$) n = 14	4.3	4.9	6.7	6.3	6.8	6.4	6.3	4.7
T-P n = 8	3.4	4.0	6.6	6.0	7.0	6.1	6.0	4.5
CON. n = 6	5.5	6.2	5.8	5.7	6.5	6.7	6.7	5.0
SITE 2 ($\bar{x}$) n = 15	4.2	5.3	6.3	6.1	6.7	6.5	6.3	5.2
T-P n = 18	4.5	5.3	6.5	6.4	6.9	6.9	6.5	5.3
CON. n = 18	3.9	5.4	6.1	5.9	6.4	6.1	6.1	5.1
SITE 3 ($\bar{x}$) n = 16	3.9	5.2	6.5	6.1	6.7	6.4	6.8	5.1
T-P n = 8	4.0	5.0	6.4	6.4	6.6	6.9	7.0	4.9
CON. n = 8	3.8	5.4	6.6	5.8	6.8	5.9	6.5	5.3
OVERALL CATEGORY $\bar{x}$	4.1	5.2	6.5	6.2	6.7	6.4	6.5	5.0

RATING SCALE

7 = COMPLETELY IMPORTANT
6 = IMPORTANT
5 = SLIGHTLY IMPORTANT
4 = NEITHER IMPORTANT NOR UNIMPORTANT
3 = SLIGHTLY UNIMPORTANT
2 = UNIMPORTANT
1 = COMPLETELY UNIMPORTANT

Table X. Mean Ratings of the Importance of the CPRS categories in Discriminating a Good Consultant.

C P R S C A T E G O R I E S

	APPEARANCE	EXPRESSION	RELATIONSHIP	CONCERN	KNOWLEDGE	PERCEPTION	INTERPRE-TATION	TERMINATION
SITE 1 ($\bar{x}$)	7	6	2	4	1	3	5	8
T-Ps	7	6	2	3	1	4	5	8
CON.	7	6	2	5	1	3	4	8
SITE 2 ($\bar{x}$)	8	6	4	5	1	2.5	2.5	7
T-Ps	8	6	1	5	2	3	4	7
CON.	8	6	4	5	1	3	2	7
SITE 3 ($\bar{x}$)	8	6	2	5	1	4	3	7
T-Ps	8	6	4	5	2	1	3	7
CON.	8	6	2	5	1	4	3	7
OVERALL	8	6	2	5	1	3.5	3.5	7

Ranking: "1" = Most important category in discriminating a good consultant.

"8" = Least important category in discriminating a good consultant.

Table XI. Average Ranking of the CPRS categories regarding their order of Importance in Discriminating a Good Consultant.

consulting contacts with any of the consultant-trainees. From Site #3, three judges had consulting experience with two of the consultant-trainees.

- - - - - - - - - - - - - - - - -
Insert Table XII about here
- - - - - - - - - - - - - - - - -

	PERSONAL CONTACTS					CONSULTING CONTACTS			
	Trainee 1	Trainee 2	Trainee 3	Trainee 4		Trainee 1	Trainee 2	Trainee 3	Trainee 4
SITE 1					SITE 1				
TPs	0	0	2	0	TPs	0	0	0	0
CON	0	2	0	2	CON	0	0	0	0
TOTAL	0	2	2	2	TOTAL	0	0	0	0
SITE 2					SITE 2				
TPs	0	0	0	0	TPs	0	0	0	0
CON	0	0	0	0	CON	0	0	0	0
TOTAL	0	0	0	0	TOTAL	0	0	0	0
SITE 3					SITE 3				
TPs	0	2	8	6	TPs	0	0	0	0
CON	5	7	6	8	CON	3	0	1	0
TOTAL	5	9	14	14	TOTAL	3	0	1	0

Table XII. Number of Judges from each Site who were familiar with the Consultant-Trainees they rated.

DISCUSSION

The results of this study socially validated the importance of effects of a consultant training program. In addition, the study demonstrated that social validity ratings can differ significantly due to a judge's program affiliation and the individual being rated.

Regarding the social validation of the effects of the consultant training program, Kendall correlation results support the hypothesis that consultant-trainees who scored higher objectively also received more favorable subjective ratings by the relevant judges. These results seem to validate the importance of the consultant training program's effects by demonstrating that the objective behavioral changes measured in the Smart (1980) study are perceived subjectively as improvements by relevant consumers.

Since the analysis of variance, however, did not reveal any significant differences in judges' ratings of pre- and post-training tapes, one might argue that the training program did not make a significant difference in how the consultant-trainees were perceived. One problem with this conclusion, however, lies in the potential confound that Trainee #4 demonstrated no behavioral change as measured objectively from pre- to post-training (53% both times). Examining the significant interaction effect between individual consultant-trainee and the level of training also may aid in dismissing this conclusion. As depicted in Figure 7 Trainees #1 and #2 received lower pre-ratings that post, whereas Trainees #3 and #4 received higher pre-ratings that post. A closer look at this Figure also demonstrates that Trainees #1 and #2 received pre-ratings that were less

than "acceptable" on the CPRS (2.92 and 2.77 respectively) whereas Trainees #3 and #4 received pre-ratings well above the 3.0 level of acceptability (3.93 and 4.03 respectively). These results may indicate that consultant-trainees who have less than acceptable consultant skill levels before training benefit more from the consultant training program and are perceived as making greater improvements, whereas consultant-trainees who have acceptable skill levels prior to the workshop do not gain as much from the training program. These results, may indicate that the consultant-training program should be refined to meet the needs of more skilled consultants.

The investigation of how several variables (e.g., judges' program affiliation, judges' position, individual consultant-trainee, and consultant-trainee's level of training) influence social validity ratings revealed some interesting findings. Consistent with Trotzer (1976), the analysis of variance reflected a significant main effect for the judge's program affiliation (site). This main effect is qualified by a significant interaction effect between site and individual consultant-trainee. Inspection of this interaction in Figure 5 reveals that each site ranked the individual trainees' competence the same, with one exception. In other words, all three sites rated Trainee #3 as the best, Trainee #4 next, Trainee #1 third, and Trainee #2 last (with the exception that Site #1 rated Trainee #4 first and Trainee #3 second). Further analysis of this interaction shows that in all cases Site #2 gave the lowest ratings and Site #3 rated the highest, with Site #1 giving ratings between the other two sites. The only exception to this was that Site #3 rated Trainee #2 lower than Site #1.

Identifying the major reasons for the significant differences in ratings between sites is complex. An analysis of covariance tested the hypothesis that how one defined an "ideal" consultant may differ from site to site and consequently contribute to the different ratings. The analysis of covariance results, however, revealed that the ratings of the "ideal" consultant did not contribute significantly to these differences. As suggested by Kazdin (1977), there is a strong possibility that these site differences are at least partially due to some judges being more familiar than others with the viewed consultant-trainees. Comparing Figure 3 and Table XII reveals that, in general, as the amount of personal familiarity increases, the mean subjective rating increases. Site #2, for example, gave the lowest mean rating of 2.97 and none of the judges from Site #2 were familiar with any of the consultant-trainees. Site #1 gave an overall rating of 3.58, and although most of the judges were unfamiliar with the consultant-trainees, two had met Trainee #2, two had met Trainee #3, and two had met Trainee #4. Site #3 gave the highest mean rating of 3.79, and this corresponded with a high degree of personal familiarity (e.g., 5 judges knew Trainee #1, 9 knew Trainee #2 and 14 knew Trainees #3 and #4). These findings suggest that as judges become more familiar with the individual they are rating, previous perceptions and interactions may influence their ratings. Further analyses, such as an analysis of covariance with personal familiarity as the covariate, would be helpful in confirming this hypothesis. A multiple comparison test to followup the analysis of variance would also be helpful in identifying exactly where the significant differences lie.

The analysis of variance also indicated a significant main effect

for the individual consultant-trainee factor. This finding supports findings from several counselor studies (Dell and Schmidt, 1976; Peoples and Dell, 1975; Trotzer, 1976). This effect may be partially a result of the differences in each consultant-trainee's objective behavior (Smart, 1980). As revealed by Kendall's tau and shown in Figure 9, in general, as a consultant-trainee's behavioral performance improved, the consultant-trainee was subjectively perceived more favorably. This explanation is consistent with the findings that counselor behaviors significantly influence judges' ratings (Siegel and Sell, 1978; Haase and Tepper, 1972; Scheid, 1976).

Consistent with previous findings (Dell and Schmidt, 1976; Schmidt and Strong, 1970; Davis, 1980), the consultant-trainee's level of training (pre- or post-training) did not produce a significant effect upon judges' ratings. There were two significant interaction effects involving the consultant-trainee's level of training, as depicted in Figures 7 and 8. The interaction effect involving the individual consultant-trainees and their level of training was discussed earlier as it related to the social validity of the consultant training program. The significant interaction between site and consultant-trainee's level of training (Figure 8) reveals that both Site #1 and Site #2 rated pre-tapes slightly higher than post-tapes, whereas Site #3 rated post-tapes higher than pre-tapes. One possible explanation for this discrepancy is that the consultant training program was specifically developed for Site #3, and consequently, in the eyes of its consumers, the training may have produced more beneficial results from this Site.

The lack of a significant main effect due to the judges' position was somewhat surprising since most evidence in the counselor literature supports the significant influence of this variable (Bishop, 1971; Brown and Cannaday, 1969; LaCrosse, 1977). Trotzer (1976), however, also found no significant differences between ratings according to judges' position. Thus at this time, it is difficult to speculate about the influence of this variable. There was a significant interaction found between judges' position and individual consultant-trainee, as shown in Figure 6. Close examination of this figure reveals that both Teaching-Parents and consultant judges rated the consultant-trainees in the following order from lowest to highest: Trainee #2, Trainee #1, Trainee #4 and Trainee #3. The interaction effect appears to stem from the consultant judges being more discriminating in their ratings, since consultants gave both the highest and the lowest mean ratings. One possible explanation for the greater variability in consultants' ratings is that consultants were more experienced in the Teaching-Family Model than the Teaching-Parents (mean = 46.5 months for consultants and 8.1 months for Teaching-Parents). This additional experience in the Teaching-Family Model may have enabled consultant judges to discern more subtle differences in consultant-trainee behavior.

The current research attempted to improve the social validity methodology by incorporating an established, reliable, and valid measurement scale, the CPRS. (See Table V.) Although the CPRS was developed for use with a counselor population, the results suggest that the CPRS categories are useful and valid for measuring a consultant population. The judges, for example, gave a mean overall rating of 4.6 for an ideal consultant

using the CPRS. This approaches a "perfect" 5.0 rating, suggesting that the CPRS is an appropriate and valid measure of desirable consultant qualities. The eight CPRS categories also were rated and ranked by the judges according to their importance in discriminating a good consultant. The results in Table X reveal that 5 of the 8 categories were rated above 6.0 ("Important") 2 were rated 5.0 or above ("Slightly Important"), and only 1 was rated 4.1 ("Neither Important nor Unimportant"). The overall mean rating for the CPRS, collapsing all categories, was 5.83, which approaches a 6.0 rating of "Important".

Table XI reflects the average rankings of each CPRS category with "1" being the most important category in discriminating a good consultant. This table shows the close agreement between Teaching-Parent and consultant judges at each of the three sites concerning the importance of the categories. The results of the CPRS ratings and rankings again suggest the validity and reliability of this measurement scale with this consultant population. Further statistical analyses would be helpful in confirming the reliability and validity of the CPRS with this population. These analyses might include correlating the degree of agreement among the judges with the ratings and rankings of the CPRS categories or with the ratings of an ideal consultant.

In summary, this research socially validates the effects of the consultant training program described by Smart (1980) and demonstrates the influence of several variables on social validity ratings. Specifically, the results confirmed that two variables, judges' program affiliation and an individual factor for the person being rated, can significantly influence a judges' ratings. The variables of the judge's position and the level of

training of the individual being rated did not produce significant differences in ratings. Future social validity research should consider these findings in their design and choice of relevant judges in order to avoid potential confounds in their results.

Further social validation research would be beneficial in two areas. First, additional studies are needed that investigate how other judge and client variables influence ratings. Specific variables warranting further investigation may include sex of judge and client and judges' position as a potential influence. These variables are suggested since previous results have been conflicting. Second, continuing to refine the measurement instruments used in social validation studies will help broaden the scope of future research. In particular, psychometric assessment of complex variables such as convergent and discriminant validity, test-retest reliability, and the potential of set responses will be beneficial in assuring that the results of social validation studies are true and valid, not merely artifacts of the measurement system employed.

In conclusion it should be stated that if the field of applied behavior analysis is to realize its importance in the study of human behavior it must continue its new emphasis on subjective evaluation.

REFERENCES

Azrin, N.H. & Armstrong, P.M. The "mini-meal" -- a method for teaching eating skills to the profoundly retarded. Mental Retardation, 1973, 11-13.

Bishop, J.B. Another look at counselor, client and supervisor ratings of counselor effectiveness. Counselor Education and Supervision, 1971, 10(4), 319-323.

Bornstein, P.H., Bach, P.J., McFall, M.E., Friman, P.C. & Lyons, P.D. Application of a social skills training program in the modification of interpersonal deficits among retarded adults: a clinical replication. Journal of Applied Behavior Analysis, 1980, 13, 171-176.

Braukmann, C.J., Kirigin, K.A., & Wolf, M.M. Achievement Place: The Researchers' perspective. Paper presented at the meeting of the American Psychological Association, Washington, D.C., September, 1976.

Briscoe, R.V., Hoffman, D.B. & Bailey, J.S. Behavioral community psychology: Training a community board to problem solve. Journal of Applied Behavior Analysis, 1975, 8, 157-168.

Brown, D. & Cannaday, M. Counselor, counselee, and supervisor ratings of counselor effectiveness. Counselor Education and Supervision, 1969, 8, 111-118.

Claiborn, C.D., & Schmidt, L.D. Effects of presession information on the perception of the counselor in an interview. Journal of Counseling Psychology, 1977, 24, 4, 259-263.

Dancer, D.D., Braukmann, C.J., Schumaker, J.B., Kirigin, K.A., Willner, A.G., & Wolf, M.M. Training and validation of behavior observation and description skills. Behavior Modification, 1978, 2, 1, 113-134.

Davis, M.A. Some procedures for developing training programs in large scale natural settings. Dissertation (Ph.D.), Lawrence, Kansas: University of Kansas, Human Development Dept., 1980.

Dell, D.M., & Schmidt, L.D. Behavioral cues to counselor expertness. Journal of Counseling Psychology, 1976, 23, 197-201.

Engelkes, J.R., & Roberts, R.R. Rehabilitation counselor's level of training and job performance. Journal of Counseling Psychology, 1970, Vol. 17(6), 522-526.

Farina, A., Allen, J.G., & Saul, B.B. The role of the stigmatized in affecting social relationships. Journal of Personality, 1968, 36, 169-182.

Farina, A., Gliha, D., Boudreau, L.A., Allen, J.G., & Sherman, M. Mental illness and the impact of believing others know about it. Journal of Abnormal Psychology, 1971, 77, 1-5.

Fawcett, S.B., & Miller, L.K. Training public-speaking behavior: an experimental analysis and social validation. Journal of Applied Behavior Analysis, 1975, 8, 125-136.

Foxx, R.M. & Azrin, N.A. Restitution: A method of eliminating aggressive-disruptive behavior of retarded and brain damaged patients. Behavior Research and Therapy, 1972, 10, 15-27.

Foxx, R.M. & Shapiro, S.T. The timeout ribbon: A nonexclusionary timeout procedure. Journal of Applied Behavior Analysis, 1978, 11, 125-136.

Gruber, B., Reeser, R., & Reid, D.J. Providing a less restrictive environment for profoundly retarded persons by teaching independent walking skills. Journal of Applied Behavior Analysis, 1979, 12, 285-297.

Haase, R.F., & Tepper, D.T., Jr. Nonverbal components of empathic communication. Journal of Counseling Psychology, 1972, 19, 5, 417-424.

Hartley, D.L. Perceived counselor credibility as a function of the effects of counseling interaction. Journal of Counseling Psychology, 1969, 16, 63-68.

Hepner, P.P., & Pew, S. Effects of diplomas, awards, and counselor sex on perceived expertness. Journal of Counseling Psychology, 1977, 24, 2, 147-149.

Johnson, R.W., & Frederickson, R.J. Effect of financial remuneration and case description on counselor performance. Journal of Counseling Psychology, 1968, 15, 130-135.

Kazdin, A.E. Assessing clinical or applied importance of behavior change through social validation. Behavior Modification, 1977, 1, 4, 427-452.

Kazdin, A.E. Acceptability of alternative treatments for deviant child behavior, Journal of Applied Behavior Analysis, 1980, 13, 259-273.

Kelz, J.S. The development and evaluation of a measure of counselor effectiveness, Personnel and Guidance Journal, 1966, 44, 511-516.

Kent, R.N., & O'Leary, K.D. A controlled evaluation of behavior modification with conduct problem children. Journal of Consulting and Clinical Psychology, 1976, 44, 586-596.

LaCrosse, M.B. Comparative perceptions of counselor behavior: A replication and extension. Journal of Counseling Psychology, 1977, 24, 6, 464-471.

Maloney, D.M., Harper, T.M., Braukmann, C.J., Fixsen, D.L., Phillips, E.L., & Wolf, M.M. Effects of training predelinquent girls on conversation and posture behaviors by teaching-parents and juvenile peers. Journal of Applied Behavior Analysis, 1976, 9, 371.

Maloney, K.B., & Hopkins, B.L. The modification of sentence structure and its relationship to subjective judgments of creativity in writing. Journal of Applied Behavior Analysis, 1973, 6, 425-434.

Martin, R. Legal challenges to behavior modification: Trends in schools, corrections and mental health. Champaign, Illinois: Research Press, 1975.

McFall, R.M. & Twentyman, C.T. Four experiments on the relative contribution of rehearsal, modeling, and coaching to assertion training. Journal of Abnormal Psychology, 1973, 81, 199-218.

Minkin, N., Braukmann, C.J., Minkin, B.L., Timbers, G.D., Timbers, B.J., Fixsen, D.L., Phillips, E.L., & Wolf, M.M. The social validation and training of conversational skills. Journal of Applied Behavior Analysis, 1976, 9, 127-139.

Myers, J.L. Fundamentals of Experimental Design. Boston: Allyn and Bacon, 1967.

O'Connor, R.D. Relative efficacy of modeling, shaping, and the combined procedures for modification of social withdrawal. Journal of Abnormal Psychology, 1972, 79, 327-334.

Patterson, G.R. Interventions for boys with conduct problems: Multiple settings, treatments and criteria. Journal of Consulting and Clinical Psychology, 1974, 42, 471-481.

Patterson, L.E. Counselor education as a process of adult socialization, Dissertation Abstracts, Volume XXVI, Number 11, 1966.

Peoples, V.Y., & Dell, D.M. Black and white student preferences for counselor roles. Journal of Counseling Psychology, 1975, 22 (6), 529-534.

Porterfield, J.K., Herbert-jackson, E., & Risley, R.R. Contingent observation: An effective and acceptable procedure for reducing disruptive behavior of young children in a group setting. Journal of Applied Behavior Analysis, 1976, 9, 55-64.

Ryan, C.W., Johnson, E.G., Folsom, C.J., & Cook, K.E. The evaluation of an instrument to measure counselor effectiveness. Measurement and Evaluation in Guidance, 1970, 3, 119-124.

Scheid, A.G. Clients' perception of the counselor: The influence of counselor introduction and behavior. Journal of Counseling Psychology, 1976, 23, 6, 503-508.

Schmidt, L.D., & Strong, S.R. "Expert" and "inexpert" counselors. Journal of Counseling Psychology, 1970, 17, 115-118.

Siegel, J.D., & Sell, J.M. Effects of objective evidence of expertness and nonverbal behavior on client-perceived expertness. Journal of Counseling Psychology, 1978, 25, 3, 188-192.

Smart, D.A. The development and evaluation of a training program for residential child care worker consultants. Thesis (M.A.) Omaha, Nebraska: University of Nebraska at Omaha, Psychology Department, 1980.

Thomas, K.R., & Britton, J.O. Correlates of counseling effectiveness in employment interviewers. Journal of Employment Counseling, 1971, 8(2), 91-95.

Trotzer, J.P. Perceptions of counselor performance: A comparative study. Counselor Education and Supervision, 1976, 16(2), 126-134.

Walker, H.M. & Hops, H. Use of normative peer data as a standard for evaluating classroom treatment effects. Journal of Applied Behavior Analysis, 1976, 9, 159-168.

Werner, J.S., Minkin, N., Minkin, B.L., Fixsen, D.L., Phillips, E.L., & Wolf, M.M. Intervention package: An analysis to prepare juvenile delinquents for encounters with police officers. Criminal Justice and Behavior, 1975, 2, 55-83.

Wike, E.L. Data Analysis: A Statistical Primer for Psychology Students. Chicago, Illinois: Aldine-Atherton, Inc., 1971.

Willner, A.G., Braukmann, C.J., Kirigin, K.A., Fixsen, D.L., Phillips, E.L., & Wolf, M.M. The training and validation of youth-preferred social behaviors of child-care personnel. Journal of Applied Behavior Analysis, 1977, 10, 219-230.

Wolf, M.M. Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. Journal of Applied Behavior Analysis, 1978, 11, 203-214.

APPENDIX A

Participant Background Information Form

and

Informed Consent Forms

Today's Date:_____

"Social Validation of Group Home Consultant Skills" Study

PARTICIPANT BACKGROUND INFORMATION FORM

I. Participant Information

1. Name:_____
2. Address:_____
   _____
3. Phone: AC ( ) -_____
4. Age:_____
5. Sex: [ ] M [ ] F
6. S.S.#:_____
7. T-F Site affiliated with:
   _____

II. Educational Background

1. [ ] High School Diploma
2. [ ] Some college, but no degree attained
3. College History:

Degree(s) attained	Major field(s) of study

III. Teaching-Family Employment History

Position(s) held (job title)	Dates began & ended (mo/yr)	Site affiliated with	Major responsibilities (outline briefly)	Specific training received for this position (if any)

*NOTE: If you are or were a TP, please indicate whether or not you are certified. [ ] Yes [ ] No

IV. Specific Teaching-Family Consultation Experience

A. Consultation services you have RECEIVED: (*applicable to TPs and assistant TPs)

Type of service received	Dates began & ended (mo/yr)	Average frequency (# times/month)
Yes No 1. Telephone Consult. [ ] [ ]	_____to_____	_____/month
2. In-home visits [ ] [ ]	_____to_____	_____/month
3. Crisis intervention[ ] [ ]	_____to_____	_____/month
4. Treatment Planning [ ] [ ]	_____to_____	_____/month
5. Other (please specify):		

B. Consultation services you have PROVIDED for TPs: (*applicable to consultants)

Type of service provided	# of TP couples you've provided this service to	Total amount of experience with this activity (# of months)
Yes No 1. Telephone Consult. [ ] [ ]	_____ couples	_____ months
2. In-home visits [ ] [ ]	_____ couples	_____ months
3. Crisis intervention [ ] [ ]	_____ couples	_____ months
4. Treatment planning [ ] [ ]	_____ couples	_____ months
5. Other (please specify):		

C. Consultant training you have RECEIVED: (*applicable to consultants)

Type of consultant activity trained in (in-home visits, telephone, etc.)	Format of training (workshop, discussions, etc.)	Length of training (days)

D. Consultant training you have GIVEN: (*applicable to consultant trainers)

Type of consultant activity trained (in-home visits, telephone, etc.)	Format of training (workshop, etc.)	Length of training (days)	# times you've given this training	Total # consultants you've trained

(Relevant Judge's Form)

CENTER FOR THE STUDY OF YOUTH DEVELOPMENT
Boys Town, Ne. 68010

INFORMED CONSENT TO PARTICIPATE IN THE
"SOCIAL VALIDATION OF GROUP HOME CONSULTANT
SKILLS" STUDY

This form represents the voluntary consent of _____ (First Name) _____ (Middle Initial) _____ (Last Name) to participate in the project known as "Social Validation of Group Home Consultant Skills" study.

PURPOSE: I understand that the purpose of the study is to assess the social validity of the skills taught during the Teaching-Family Consultation Workshop and to assess any differences in validation ratings among various groups of relevant judges.

PROCEDURES: I understand that my consent allows the collection, analysis, and dissemination of information regarding my ratings of videotapes of consultant trainee interactions with teaching-parents and youths. I understand consent also allows collection, analysis, and dissemination of background information concerning my educational and occupational experiences.

DISCOMFORTS OR RISKS: I understand that there is no physical risk or discomfort involved. I understand that the potential social or psychological risk I may experience is that my name may become publicly associated with particular research results. I understand that I am protected from this risk by safeguards described in the PRIVACY PROTECTION section of this form.

PRIVACY PROTECTION: I understand that every effort will be made to insure confidentiality and my anonymity. The Study Director of the research will keep information about me in locked files. I understand that publication of study results in any form will protect my privacy and disguise me by using pseudonyms or code numbers instead of my real name. Also, information about me may be combined with information about others in a group fashion to further protect my privacy.

BENEFITS: I understand that the potential benefits of the study are that it will provide information concerning which consultant skills are preferred by relevant consumers and professional colleagues. This information can then be used to improve the Teaching-Family Consultation Workshop and to improve the methodology of social validation studies. The study also may permit distribution of the Consultation Workshop methods to others.

ALTERNATIVES: I understand that the alternative to my participation in the research is to not participate.

--Please read both sides--

RIGHT TO WITHDRAW: I understand that if I wish to withdraw my consent later I may freely do so without penalty even after I sign this consent form. I agree that such a withdrawal will be made in writing to the Study Director of the research.

RIGHT TO INQUIRY: I understand that I have the right to inquire at any time about the procedures described in this document. I understand that I can direct these inquiries to the Study Director in writing or by calling collect at (402) 498-1550 or 498-1990.

I understand that my signature below signifies my voluntary informed consent for my participation in the research study.

_____ _____

(Participant Signature) (Date)

Inquiries may be sent to:

Dianne Smart, B.A., Study Director (402) 498-1550
Karen B. Maloney, Ph.D. (402) 498-1990
Center for the Study of Youth Development
Boys Town, NE 68010

(Consultant Trainee's Form)

INFORMED CONSENT TO PARTICIPATE IN THE
"SOCIAL VALIDATION OF GROUP HOME CONSULTANT
SKILLS" STUDY

This form represents the voluntary consent of _____ (First Name) _____ (Middle Initial) _____ (Last Name) to participate in the project known as "Social Validation of Group Home Consultant Skills" study.

PURPOSE: I understand that the purpose of the study is to assess the social validity of the skills taught during the Teaching-Family Consultation Workshop and to assess any differences in validation ratings among various groups of relevant judges.

PROCEDURES: I understand that my consent allows the collection, analysis and dissemination of information regarding ratings of my pre and post Consultation Workshop videotapes. I understand that some teaching-family personnel at Teaching-Family sites all over the United States and some non-teaching-family child care agency personnel in the Omaha area will be shown selected pre and post workshop videotapes in which I participated. I understand ratings on various dimensions (e.g., satisfaction, pleasantness, effectiveness, etc.) will be obtained from these personnel. These ratings will be compared to other measures derived from the tapes.

DISCOMFORTS OR RISKS: I understand that there is no physical risk or discomfort involved. I understand that the potential social or psychological risk I may experience is that my name may become publicly associated with particular research results and that other child care professionals will view videotapes in which I appear. I understand that I am protected from these risks by safeguards described in the PRIVACY PROTECTION section of this form.

PRIVACY PROTECTION: I understand that every effort will be undertaken to insure confidentiality and my anonymity. The Study Director of the research will keep information about me in locked files. I understand that publication of study results in any form will protect my privacy and disguise me by using pseudonyms or code numbers instead of my real name. Also, information about me may be combined with information about others in a group fashion to further protect my privacy.

BENEFITS: I understand that the potential benefits of the study are that it will provide information concerning which consultant skills are preferred by relevant consumers and professional colleagues. This information can then be used to improve the Teaching-Family Consultation Workshop and to improve the methodology of social validation studies. The study also may permit distribution of the Consultation Workshop methods to others.

--Please read both sides--

ALTERNATIVES: I understand that the alternative to my participation in the research is to not participate.

RIGHT TO WITHDRAW: I understand that if I wish to withdraw my consent later I may freely do so without penalty to me even after I sign this consent form. I agree that such a withdrawal will be made in writing to the Study Director of the research.

RIGHT TO INQUIRY: I understand that I have the right to inquire at any time about the procedures described in this document. I understand that I can direct these inquiries to the Study Director in writing or by calling collect at (402) 498-1550 or 498-1990.

I understand that my signature below signifies my voluntary informed consent for my participation in the research study.

_____ _____

(Participant Signature) (Date)

Inquiries may be sent to:

Dianne Smart, B.A., Study Director (402) 498-1550
Karen B. Maloney, Ph.D. (402) 498-1990
Center for the Study of Youth Development
Boys Town, NE 68010

APPENDIX B

Standardized Instructions Used During Implementation of the Study

"Social Validation of Group Home Consultant Skills" Study

STANDARDIZED INSTRUCTIONS

I. INTRODUCTION:

1. Thank everyone for agreeing to participate in the study.

2. Name of Study: "Social Validation of Group Home Consultant Skills"

   Purpose: To assess the social validity of the skills taught in the Teaching-Family Consultation Workshop and to assess any differences in validation ratings among various groups of relevant judges (participants).

3. Informed Consent/Background Info Forms

   Collect informed consent and background info forms. (If lost or forgotten, get informed consent now. Can wait for background information.)

II. GENERAL PROCEDURES:

You will be shown 8 different videotaped interactions between a consultant and a Teaching-Parent. These interactions involve a problem discussion concerning a youth in the Teaching-Parent's home.

Before we actually view and rate the tapes, I would like for you to think about your conception of how an "ideal" consultant would discuss a youth problem with a Teaching-Parent. Then I'd like for you to rate this "ideal" consultant on the rating scale and to use this ideal as a reference for your ratings of each of the videotaped consultants.

Summary: So we'll rate the "ideal" consultant, then the videotaped consultants. After completing the ratings and I've collected them, I'll be happy to answer questions and to give any further explanation of what the study is all about.

III. EXPLAIN RATING SCALE:

Now before we begin rating our ideal and then rating the videotaped consultants, I would like to familiarize you with the rating scale. (Hand out definition sheet and scale to rate "IDEAL" consultant.)

RATING SCALE

1. Identify the 8 major catogories (ALL CAPS).

2. Each category is rated on a 5-pt. scale (rate by circling).

1	2	3	4	5
(unsatisfactory)		(acceptable)		(outstanding)

(NOTE: 3 - acceptable) -- Stress how this differs from midpoint on traditional TF 7-pt. scale

3. Each major category has several subdivisions which should be considered before making your rating.

(NOTE: Each subdivision is NOT rated; it is only used to help make your decision for rating the overall category.)

4. Each subdivision is clarified by both a favorable (on right) and unfavorable (on left) descriptive statement.

5. You should position your rating scale in the manual so the category titles are aligned horizontally (i.e., lines match up). This will allow quick, easy access to the definitions of the subdivisions. Be sure to put the rating scale on the page opposite from the categories you're rating (so the arrows pt. to the page with the definitions). So if rating appearance through communication, put rating scale on right hand page of manual. If rating knowledge through termination, put rating scale on left hand page of manual.

6. Give examples - using APPEARANCE & KNOWLEDGE.

7. Remind participants that they should rate only the 8 major categories and they should rate by circling the number. (Be sure to circle only numbers, not "in-between.")

IV. RATE "IDEAL" CONSULTANT:

1. Ask participants to think of their conception of how an "ideal" consultant would discuss a youth problem with a TP.

2. Make sure have form with their code # and for "IDEAL" consultant.

3. First rate this "ideal" consultant on all 8 major categories by circling the # which represents your best estimate of how and ideal consultant would fulfill this category.

4. On back of rating form, you are asked to list the 5 most important skills or personal characteristics that an "ideal" consultant should possess in a youth problem discussion with a TP. Also please list the 5 most important behaviors or personal characteristics that an "ideal" consultant should avoid in this situation. Please be brief and take no longer than 5 minutes to list these things.

5. NOTE: This is first time participants read scale closely. If they have questions about it, definitions, etc., please answer.

6. Collect forms and make sure properly completed and labelled (or may want to keep them as reference).

V. GENERAL INSTRUCTIONS FOR RATING TAPES

1. Hand out rating forms for viewing tapes. Each person should have 8 forms, each already labelled with their observer code #. (Use the code key list in handing out the forms.)

2. Explain that now will view 8 different interactions between a consultant and a TP. You will rate each consultant using the same rating scale as you used for the "ideal" rating. IMPORTANT! Please use the same criteria or standard as you did to rate an "ideal" consultant when rating each of these consultants.

3. Also on the back of the rating sheet you are asked 2 Familiarity questions just to find out if you are familiar with the person and how well you know them.

4. Summary: So you should complete the rating scale and familiarity questions for each consultant you view. Please fill in the consultant code # that I will give you for each one.

5. SETTING:

   In order to help you with the setting, here's a brief description of scene: The consultant will begin the interaction by asking the TP if they have any problems they would like to discuss. The TP will being up one problem concerning a youth in their home. The problem is the youth is shy and unresponsive (or stubborn and moody). A problem discussion follows. Any questions about the scene?

   These video segments range in length from approximately 3 to 10 minutes. After each segment, you will be given time (max = 5 min.) to rate that particular consultant.

6. CLARIFICATIONS:

   A. This situation you will view is actually only half of entire role-play sequence -- 1 problem discussion, when entire conversation included 2 problems. Therefore, when rating TERMINATION, do not expect to see final departure statements.

B. Consultants may suggest use of the "structured" system in their discussion. Explain this is comparable to a point system, with positive and negative fx being comparable to giving and taking points.

7. Everyone please rate carefully and honestly, think of and rate each category individually. Please do not talk, etc., so as to influence others. You may go back and change ratings after viewing another tape if you wish, but if you do, just "X" your first rating and circle another rather than erase. (If you want to change ratings before view another tape, that's fine to erase.)

   Any questions about what is to proceed? (Because once start rating tapes will not be able to answer questions until finished.)

VI. BEGIN RATING TAPES:

1. Remond them to use same criteria as did to rate "Ideal" consultant.

2. Begin tape and identify consultant (consultant always on left side of screen). Stop after each consultant and allow 5 minutes (max) for rating.

3. *Be sure to give participants consultant # to fill in on their rating sheets (see attached page).

4. Continue until all consultants (8) are viewed.

5. Take a stretch break halfway through??

6. After completed, collect all forms and check each person's to make sure properly labelled and completed. Put in appropriate folders.

VII. RANK/RATE QUESTIONS ON SCALE:

Now, one final request before we finish.

Would like you to rank and rate each of the 8 major categories on the rating scale concerning how important each category is in determining/discriminating a good consultant in a problem discussion situation (such as portrayed in the tapes).

Again, please rate/rank only the overall categories, but be sure to take into consideration the subcategories.

(Hand out forms marked "RATE/RANK SCALE" -- make sure code #s match participants.)

Explain procedure:

A. Read instructions on back side.

B. First rank each major category (in all CAPS -- e.g., appearance, expressions, etc.) from 1 to 8 in the ranking box ☐ . Let 1 indicate which you feel to be the most important category in determining a good consultant in a problem discussion situation and 8 be the least important. Give example: Appearance = most important = 1; knowledge = least important = 8. Questions? *Give time to rank now, before rating so not confusing.

C. Next, rate each major category on a 7-pt. scale concerning how important this particular category is in determining a good consultant in a problem discussion situation. Rate by circling your best estimate beside the major category.

Rating Scale

7 = Completely Important
6 = Important
5 = Slightly Important
4 = Neither Important nor Unimportant
3 = Slightly Unimportant
2 = Unimportant
1 = Completely Unimportant

Give example: So if feel that APPEARANCE is completely important, circle "7". If feel that it is completely unimportant, circle "1".

Questions?

Give time to rate.

D. Collect all forms and check to make sure properly labelled and completed.

VIII. FINAL PROCEDURES:

1. Open the floor to questions concerning the study or tapes.

2. Thank everyone.

3. (Explain how will get $5.)

APPENDIX C

Checklist of Behaviors for Simulation #4: Part A

(This checklist was used in a previous study (Smart, 1980) to determine the objective scores for the viewed videotaped consultant-trainees in the current study. The objective score was the percentage of appropriate consultant behaviors displayed by the consultant-trainees.)

SIMULATION FOUR: PROBLEM DISCUSSION AND FEEDBACK TO A CONSULTEE

PART I

List of Appropriate Behaviors

____ 1. Initial praise statements.

____ 2. Initial empathy/support statements.

____ 3. Initial courtesy statements.

____ 4. Obtains a behavioral description of problem.

____ 5. Asks Teaching-Parent to be more specific.

____ 6. Mentions why Teaching-Parent should use behavioral descriptions when describing problems.

____ 7. Provides a rationale for working on the problem which includes benefits to youth, Teaching-Parents or the home.

____ 8. Asks how Teaching-Parent is presently handling the problem.

____ 9. Asks Teaching-Parent for suggestions to correct the problem.

____10. Asks for Teaching-Parent's suggestions before suggesting one himself/herself.

____11. Suggests or discusses "teaching" as a problem solution strategy.

____12. Discusses a component of effective teaching (description of behavior, use of consequence, practice or rationales).

____13. Requests acknowledgement following the solution discussion.

____14. Summarizes the solution.

____15. Suggests a check-back procedure for monitoring the problem in the future.

APPENDIX D

Table I. Objective Performance Scores for each Videotaped Consultant-Trainee Before and After Training.

	Percent Appropriate Consultant Behaviors (on Simulation #4, Part A)	
Consultant-Trainees	Pre-Training	Post-Training
Trainee 1	40%	60%
Trainee 2	20%	60%
Trainee 3	73%	87%
Trainee 4	53%	53%

	PRE		POST		COMBINED	
	Subsample (n=4)	Total Sample (n=12)	Subsample (n=4)	Total Sample (n=12)	Subsample	Total Sample
Mean	48.3	47.1	66.1	65.7	56.2	57.1
Range	21.4 - 73.3	14.3 - 73.3	53.3 - 92.9	50 - 92.9	14.3 - 92.9	21.4 - 92.9
S. D.	21.6	17.4	17.9	14.8	18.5	24.5
t-tests	n.s		n.s		n.s	

Therefore it appears subsample is statistically comparable to the total sample and can be used to make statements about the total sample.

Table II. Comparison of Objective Scores of Chosen Subsample of Consultant-Trainees with total Sample (From Smart (1980) study).

Table III. COMPARISON OF SEX OF CHOSEN SUBSAMPLE WITH TOTAL SAMPLE

	Subsample (n=4)	Total Sample (n=12)
Males	75%	75%
Females	25%	25%

APPENDIX E

TABLE IV. Results of t-test for Regular vs. Counterbalanced Viewing Order of Videotape.

							POOLED VARIANCE ESTIMATE			SEPARATE VARIANCE ESTIMATE		
	NUMBER OF CASES	MEAN	STANDARD DEVIATION	STANDARD ERROR	F VALUE	2-TAIL PROB.	T VALUE	DEGREES OF FREEDOM	2-TAIL PROB.	T VALUE	DEGREES OF FREEDOM	2-TAIL PROB.
SCORE												
REGULAR	22	3.4091	.514	.110								
					1.05	.908	-.43	44	.670	-.43	43.45	.670
COUNTERBALANCED	24	3.4734	.502	.102								

Table IV. Results of t-test for Regular vs. Counterbalanced Viewing Order of Videotaped Consultant-Trainees.

APPENDIX F

Consultant Performance Rating Scale (CPRS) and Definitions

PERSONAL FAMILIARITY WITH THIS CONSULTANT

1. Please indicate how well you know this consultant personally by checking ( ) how many contacts you've had with this person. (Contacts refer to either social or professional).

0 Contacts	1 Contacts	2-5 Contacts	5-10 Contacts	Over 10 Contacts
( )	( )	( )	( )	( )

2. Please indicate ( ) how many consulting contacts you've had with this consultant, if any. Consulting contacts refer to any contact you've had with this consultant in which you've directly experienced this person's skills as a consultant (e.g., consulted in your home; partner consultant with you, etc.)

0 Contacts	1 Contact	2-5 Contacts	5-10 Contacts	Over 10 Contacts
( )	( )	( )	( )	( )

Please describe briefly the nature of these consulting contacts:_____

_____

_____

*The CPRS was developed and copyrighted by James W. Kelz in 1961 and was originally named the Counselor Performance Rating Scale.

After observing each of the problem discussions, rate each category giving consideration to subdivisions. Circle the number which represents your best estimate of the degree to which a consultant fulfills each category.

1 (Unsatisfactory) 2 3 (Acceptable) 4 5 (Outstanding)

If your estimate is intermediate between unsatisfactory (1) and acceptable (3), circle number two (2); if your estimate is intermediate between acceptable (3) and outstanding (5), circle number four (4).

CONSULTANT PERFORMANCE RATING SCALE *

APPEARANCE	1	2	3	4	5	1	2	3	4	5	KNOWLEDGE
--Grooming											Practical knowledge--
--Poise											Application--
EXPRESSION	1	2	3	4	5	1	2	3	4	5	PERCEPTION
--Articulation											Cue alertness--
--Informal correctness											Problem identification--
RELATIONSHIP	1	2	3	4	5	1	2	3	4	5	INTERPRETATION
--Interest											Relevancy--
--Rapport											Accuracy--
--Permissiveness											Clarity--
--Empathy											Neutrality--
											Consultee involvement--
COMMUNICATION	1	2	3	4	5	1	2	3	4	5	TERMINATION
--Continuity											Summarization--
--Response appropriateness											Ease--

--Please rate both sides--

Consultant #:_____

Observer #:_____

Date:_____

UNFAVORABLE	CATEGORY	FAVORABLE
	APPEARANCE	
Appears "sloppy" and inappropriately attired		Appears well-groomed and appropriately attired
Appears tense; displays overconfidence		Appears relaxed; displays confidence
	EXPRESSION	
Mumbles; talks inaudibly; rambles in presentation		Enunciates clearly; modulates voice; expresses self concisely
Is ungrammatical in English usage		Uses correct English without seeming overly formal
	RELATIONSHIP	
Displays boredom or impatience; is condescending and patronizing		Exhibits sincere interest in consultee
Makes consultee feel ill-at-ease by unduly hurrying him; prying		Helps consultee feel at ease by beginning problem discussion in a warm, friendly way
Dominates conversation; prods consultee; tries to be subtle or shrewd		Helps consultee to talk freely about problems; responds straight-forwardly
Becomes emotionally involved in consultee's problems; lectures consultee about proper solutions for problems		Avoids passing judgment; reflects consultee's feelings; gives him opportunity to identify, analyze, and solve problems himself
	COMMUNICATION	
Interrupts; dawdles; fills pauses with idle chatter		Allows consultee to set pace; does not allow problem discussion to lose direction; uses pauses productively
Strays from subject		Makes comments consistent with what consultee is trying to convey

UNFAVORABLE	CATEGORY	FAVORABLE
	KNOWLEDGE	
Bluffs way rather than admitting ignorance		Demonstrates practical knowledge related to problem areas
Fails to apply information to consultee's problems		Utilizes information gained from previous experiences
	PERCEPTION	
Misses cues and flounders in getting consultee's meaning		Is alert in picking up cues and following through
Overemphasizes "superficial" problems; jumps to conclusions		Isolates the consultee's central problem
	INTERPRETATION	
Stresses non-pertinent information		Interprets the significant information revealed in problem discussion
Misinterprets		Interprets accurately
Confuses consultee by being vague and evasive; uses technical "jargon"		Interprets information clearly; uses vocabulary consistent with consultee's verbal ability
Makes biased or critical statements		Avoids bias and criticism; refrains from reassuring when facts are unfavorable
Does not involve consultee in assessment of self-information		Enlists consultee participation in assessing information objectively
	TERMINATION	
Allows problem discussion to extend beyond allotted time; ends abruptly		Respects time limitations; makes provisions for summary
Allows problem discussion to degenerate into casual social conversation; terminates problem discussion awkwardly		Terminates problem discussion politely and smoothly

NOTE: "Consultee" refers to the Teaching-Parent.

APPENDIX G

TABLE V. Reliability and Validity Results of Using CPRS in Several Counselor Studies

Reference	Judges Used (Number & Position)	What was Rated using CPRS	Reliability & Validity Analysis on CPRS	Results	Conclusions
Patterson, L.E. 1966	8 judges	Performance of counselor trainees in 2 videotaped counseling interviews	ANOVA reliability coefficients	Reliability of First interview r = .890 2nd interview = r = .790 Reliability estimate for combined ratings r = .769	These interjudge reliability results indicate that judges were in high agreement concerning the effectiveness of counselors as measured by Kelz's CPRS.
Kelz, 1966	3 counselor-educators 3 doctoral students 6 total judges	Performance of 30 counselor trainees in 2 videotaped counseling interviewes (Interview S & P)	1. Pearson product moment correlations of interjudge reliability	Average correlation between pairs of independent raters = .50 Averate agreement of individual judges with pooled ratings = .76	Concluded that CPRS rating scale constituted a usable, realistic measure for assessing counseling proficiency of counselor-trainees.
			2. Spearman rank-order correlation to assess validity (using rankings of counselor-trainees by practicum supervisors as criterion)	Interview S & criterion r = .42 Interview P & criteron r = .28	Low validity correlations possibly an artifact of practicum supervisor being influenced by extraneous factors (personality, class, performance, etc.)
Johnson, R.W. & Fredrickson, R.H., 1968	2 experienced counselors	Performance of 48 undergrad students (enrolled in counseling course) in a videotaped counseling interview.	Product moment correlation for interjudge reliability (for six of 8 categories on Kelz scale)	r (for individual categories) .49 - .68 ($p < .01$) Total r = .73 ($p < .01$)	High correlation reflects the reliability of the CPRS in measuring counselor effectiveness.
Ryan, Johnson, Folsom & Cook 1970	2 counselor-educators 2 doctoral candidates 4 total	Performance of 21 counselor-trainees in videotaped counseling interviews	1. Kendall coefficient of concordance (% of all 4 judges) 2. Spearman rank order correlation	Kendal W = .76 ($p < .01$) r = .76 between the 2 counselor-educators r = .75 between the 2 doctoral students	Results indicated that all judges were in high agreement as to what constitutes effective counseling as measured by Kelz's CPRS.
Thomas & Britton, 1971	2 professional persons familiar with Kelz scale	Performance of 17 employment interviews in videotaped counseling interviews	Interrater reliability coefficient	r = .74	Results further establish the reliability & usefulness of Kelz scale with a different population.

Table V. Reliability and Validity of CPRS as Reported in Several Counselor Studies

APPENDIX H

MEASUREMENT INSTRUMENT: Rating of IDEAL Consultant on CPRS

IDEAL CONSULTANT

1. Again, think of your conception of how an "ideal" consultant would discuss a youth problem with a TP.
2. List below the 5 most important skills or personal characteristics that an ideal consultant should possess or should do in this problem discussion situation.
3. Also list below the 5 most important behaviors or personal characteristics that an ideal consultant should avoid in this situation. Be specific.

Ideal consultant should possess or do:	Ideal consultant should avoid:
1.	1.
2.	2.
3.	3.
4.	4.
5.	5.

*CPRS was developed and copyrighted by James W. Kelz in 1961 and was originally named the Counselor Performance Rating Scale.

Please rate how you feel an "ideal" consultant would perform in a youth problem discussio with a TP. Rate each of the major categories on the scale, taking the subdivisions into consideration. The major categories are in all caps (e.g., APPEARANCE, KNOWLEDGE, etc.) rate by circling the number which represents your best estimate of how an ideal consultan would fulfill this category. Use the following rating scale:

1 (Unsatisfactory) 2 3 (Acceptable) 4 5 (Outstanding)

CONSULTANT PERFORMANCE RATING SCALE *

APPEARANCE	1	2	3	4	5	1	2	3	4	5	KNOWLEDGE
--Grooming											Practical knowledge--
--Poise											Application--
EXPRESSION	1	2	3	4	5	1	2	3	4	5	PERCEPTION
--Articulation											Cue alertness--
--Informal correctness											Problem identification--
RELATIONSHIP	1	2	3	4	5	1	2	3	4	5	INTERPRETATION
--Interest											Relevancy--
--Rapport											Accuracy--
--Permissiveness											Clarity--
--Empathy											Neutrality--
											Consultee involvement--
COMMUNICATION	1	2	3	4	5	1	2	3	4	5	TERMINATION
--Continuity											Summarization--
--Response appropriateness											Ease--

--Please complete both sides--

Consultant #: IDEAL

Observer #: _____

Date: _____

APPENDIX I

MEASUREMENT INSTRUMENTS:

Rating & Ranking the Consultant Performance Rating Scale

INSTRUCTIONS FOR RATING/RANKING

the Consultant Performance Rating Scale

1. Please rank and rate each major category of the Consultant Performance Rating Scale concerning its importance in discriminating a good consultant in a problem discussion situation. The major categories are in all caps (e.g., APPEARANCE, KNOWLEDGE, etc.).

2. RANK each from 1 to 8 by placing your rank in the box (e.g., Ranking: [4]). Let "1" indicate the category which you feel is most important in determining a good consultant and "8" be the least important.

3. RATE each category concerning how important this particular category is in discriminating a good consultant in a problem discussion situation. Rate by circling your best estimate beside the category (e.g., 1 2 3 4 (5) 6 7) and use the following rating scale:

   7 = Completely Important
   6 = Important
   5 = Slightly Important
   4 = Neither Important nor Unimportant
   3 = Slightly Unimportant
   2 = Unimportant
   1 = Completely Unimportant

--Over--

*The CPRS was developed and copyrighted by James W. Kelz in 1961 and was originally called the Counselor Performance Rating Scale.

(Please read instructions on other side before completing.)

CONSULTANT PERFORMANCE RATING SCALE *

APPEARANCE — Rating: 1 2 3 4 5 6 7 — Ranking: ☐
--Grooming
--Poise

KNOWLEDGE — Rating: 1 2 3 4 5 6 7 — Ranking: ☐
Practical knowledge--
Application--

EXPRESSION — Rating: 1 2 3 4 5 6 7 — Ranking: ☐
--Articulation
--Informal correctness

PERCEPTION — Rating: 1 2 3 4 5 6 7 — Ranking: ☐
Cue alertness--
Problem identification--

RELATIONSHIP — Rating: 1 2 3 4 5 6 7 — Ranking: ☐
--Interest
--Rapport
--Permissiveness
--Empathy

INTERPRETATION — Rating: 1 2 3 4 5 6 7 — Ranking: ☐
Relevancy--
Accuracy--
Clarity--
Neutrality--
Consultee involvement--

COMMUNICATION — Rating: 1 2 3 4 5 6 7 — Ranking: ☐
--Continuity
--Response appropriateness

TERMINATION — Rating: 1 2 3 4 5 6 7 — Ranking: ☐
Summarization--
Ease--

RATE/RANK SCALE

Observer #: _____

Date: _____

APPENDIX J

Table VI. Mean Ratings by each Judge for each Consultant-Trainee (Raw Data used for ANOVA)

TEACHING-PARENT RESPONSES
SITE 1

	Trainee 1		Trainee 2		Trainee 3		Trainee 4	
JUDGE	PRE	POST	PRE	POST	PRE	POST	PRE	POST
10	3.9	1.9	4.1	4.1	3.6	3.0	4.1	4.1
11	2.5	3.6	3.4	3.6	3.3	3.5	4.1	2.9
12	2.4	4.6	2.3	3.0	3.1	4.1	4.4	3.3
13	3.9	4.5	2.3	2.5	4.3	4.4	3.6	4.5
14	2.9	3.6	2.9	3.3	3.6	3.3	4.4	4.5
15	3.4	1.6	2.1	2.8	4.6	3.3	4.0	4.0
16	2.6	4.9	3.0	3.8	3.4	4.3	5.0	2.5
17	4.1	3.4	4.4	4.9	4.8	4.6	4.6	4.0
$\bar{x}$ =	3.2	3.5	3.1	3.5	3.8	3.8	4.3	3.7

CONSULTANT RESPONSES
SITE 1

	Trainee 1		Trainee 2		Trainee 3		Trainee 4	
JUDGE	PRE	POST	PRE	POST	PRE	POST	PRE	POST
18	3.9	3.5	1.4	2.0	4.9	3.8	4.9	4.8
19	3.1	4.0	3.5	2.6	4.8	4.4	4.9	3.4
20	3.3	5.0	1.4	2.4	4.3	5.0	5.0	4.8
21	3.0	4.4	3.6	2.5	3.3	3.8	3.9	2.1
22	2.8	2.5	3.9	4.1	2.6	4.0	4.5	3.9
23	2.3	4.1	2.4	1.8	3.8	1.8	4.9	2.6
$\bar{x}$ =	3.1	3.9	2.7	2.6	4.0	3.8	4.7	3.6

SITE 2

	Trainee 1		Trainee 2		Trainee 3		Trainee 4	
JUDGE	PRE	POST	PRE	POST	PRE	POST	PRE	POST
24	1.5	3.1	2.4	2.8	3.6	2.3	2.8	1.5
25	2.1	2.6	1.5	2.5	2.5	2.5	2.9	2.8
26	3.3	2.8	3.4	2.3	3.9	3.3	2.8	3.3
27	2.4	3.6	4.1	3.0	3.8	3.5	4.6	2.8
28	3.0	3.1	3.0	2.5	3.9	4.0	3.0	2.9
29	2.6	3.0	3.5	3.0	2.1	2.1	3.4	2.6
30	1.4	3.3	2.3	3.1	2.5	1.6	3.9	3.3
31	2.8	4.1	4.5	3.8	4.4	3.5	5.0	3.6
$\bar{x}$ =	2.4	3.2	3.1	2.9	3.3	2.9	3.6	2.9

SITE 2

	Trainee 1		Trainee 2		Trainee 3		Trainee 4	
JUDGE	PRE	POST	PRE	POST	PRE	POST	PRE	POST
32	2.4	2.3	2.0	2.3	3.9	2.5	1.9	2.4
33	2.4	3.1	1.4	2.3	3.6	3.4	3.5	2.8
34	2.1	3.6	1.9	1.4	3.3	3.3	2.8	2.1
35	2.8	2.5	2.5	3.0	3.4	2.8	3.6	2.4
36	2.8	3.4	2.0	3.0	4.4	4.4	4.8	4.4
37	2.0	3.4	1.3	4.0	4.0	2.9	4.0	2.3
38	3.1	1.9	3.5	2.0	4.0	3.0	3.0	2.1
39	2.5	3.3	2.0	3.8	3.3	3.5	3.6	3.6
$\bar{x}$ =	2.5	2.9	2.1	2.7	3.7	3.2	3.4	2.8

SITE 3

	Trainee 1		Trainee 2		Trainee 3		Trainee 4	
JUDGE	PRE	POST	PRE	POST	PRE	POST	PRE	POST
40	4.1	4.9	3.5	3.3	4.6	4.5	5.0	4.9
41	3.4	3.9	3.6	2.8	2.9	3.9	3.9	3.3
42	3.0	3.6	3.6	2.1	4.9	4.8	5.0	4.6
43	3.9	3.8	3.9	3.4	4.4	4.5	3.6	3.8
44	2.9	4.5	3.0	3.0	4.9	5.0	4.4	4.3
45	4.5	4.8	3.4	2.6	4.1	5.0	4.0	3.0
46	2.3	3.8	2.9	2.5	4.0	4.9	5.0	5.0
47	3.1	4.8	2.1	2.4	4.8	4.8	3.6	2.6
$\bar{x}$ =	3.4	4.3	3.3	2.8	4.3	4.7	4.3	3.9

SITE 3

	Trainee 1		Trainee 2		Trainee 3		Trainee 4	
JUDGE	PRE	POST	PRE	POST	PRE	POST	PRE	POST
48	3.5	4.1	2.3	2.5	4.6	4.6	3.8	4.0
49	2.9	4.0	1.9	3.6	5.0	5.0	3.8	4.0
50	1.9	4.3	2.9	3.6	3.9	4.3	3.6	3.3
51	3.9	4.1	2.1	2.4	4.0	4.1	3.9	3.6
52	3.5	4.3	2.1	1.8	5.0	4.5	3.6	3.4
53	2.0	4.6	1.8	2.9	4.4	4.6	4.5	4.4
54	2.9	4.9	3.5	3.9	4.8	4.9	4.8	4.8
55	3.0	4.5	2.8	3.4	3.6	4.4	5.0	4.1
$\bar{x}$ =	3.0	4.4	2.4	3.0	4.4	4.6	4.1	4.0

Table VI. Mean Ratings on CPRS by each Judge for each Consultant-Trainee (Raw Data used in ANOVA)