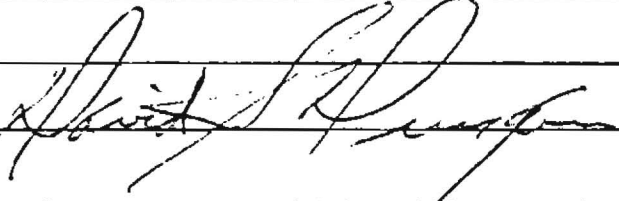AN ABSTRACT OF THE THESIS OF

Stephen Alan Jackson   for the   Master of Science

in _____ Psychology

presented on ___ July, 1982 _____ .

Title: ___ MEASURING SOCIAL ANXIETY UTILIZING ROLE-PLAY

Abstract Approved: _____

Research on behavioral assessments which utilize a role-play
analogue have produced contradictory results. The assessment of social
anxiety by such methods has been particularly laden with inconsistencies.

In the present study, twenty-two male and thirty-three female under-
graduate students were used to examine the interrelationship between two
self-report and two behavioral measures of anxiety and a measure of
social desirability. The behavioral ratings were made by two sets of
independent raters from videotapes of subjects responses to Zuroff and
Schwarz's (1978) Situation Test. The Situation Test requires subjects
to place themselves imaginatively in a series of anxiety-producing
situations described on a tape recording; their task is to respond to a
line of dialogue given on the tape. The videotapes of subject's
responses were scored using a checklist of behavioral anxiety indicators
and a global rating scale, both developed by Zuroff and Schwarz (1978).
These ratings were correlated with the two self-report measures of
anxiety. The self-report measures used in the study were the S-R
Inventory of Anxiousness (Endler, Hunt & Rosenstein, 1962) and the
Adjective Check List (Zuckerman, 1960). Subjects were also required to
fill out the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe,

1960).

Interrater reliabilities for the global ratings were moderate
(r=.49) and high for the behavioral checklist (r=.88). The results
indicated that the checklist scores were unrelated to the global ratings
or to the self-report measures of anxiety. The global ratings were also
found to be unrelated to the self-report anxiety measures. However,
the two self-report measures were found to be moderately correlated with
each other. There were no significant relationships between any of the
anxiety measures and Marlowe-Crowne social desirability.

MEASURING SOCIAL ANXIETY

UTILIZING ROLE-PLAY

———————

A Thesis

Presented to

The Department of Psychology

Emporia State University

———————
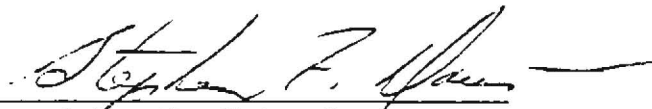
In Partial Fulfillment

of the Requirements for the Degree

Master of Science

———————

by

Stephen Alan Jackson

July, 1982

Approved for the Major Department

Approved for the Graduate Council

ACKNOWLEDGMENTS

I would like to express my appreciation to my committee members for their prompt suggestions and assistance. Appreciation is also extended to Sue Spence, Carrie Strathman, Ann Shephard, and Mary Beth Bally for their contribution to the research project.

The Emporia State Instructional Media Center was also of great assistance in conducting this study. Finally, deepest thanks are extended to my wife, Ann, and parents for their continued support and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

INTRODUCTION

This chapter is devoted to a review of the literature on the use of role-playing as an assessment technique. The review focuses on role-playing as an assessment of assertion, aggression, social skills, anxiety and fluid-rigid boundaries. In all of the studies reviewed, the experimenters created an analog situation so that the behavior to be assessed could be measured in terms of its behavioral components. Behavioral assessments were compared with self-report measures or in vivo assessment on the respective variable. In most of the studies, the role-played situation was described by the experimenter or assistants on audiotape and was designed to be as close to a real life situation as possible.

Assertion

A study investigating the reliability and validity of assertion measures was conducted by Kern and MacDonald (1980). For the purposes of this review, the aspect of their study which examined the reliability and validity of the College Women's Assertion Sample (CWAS) (MacDonald, 1978) is relevant. The CWAS is a role-play test which is used to measure or assess assertion. The authors of this study were primarily concerned with exploring any interrelatedness among various test (convergent and discriminant validity) and whether or not individual tests (such as the CWAS) are temporally stable (reliability).

In order to evaluate the property of temporal stability, 120 undergraduate women were tested and then retested at one of the following three randomly assigned retest intervals: 1 week (n=40), 4 weeks (n=40),

and 10 weeks ($\underline{n}$=40). Each individual was, thus, tested at two points in time.

In order to evaluate the properties of convergent and discriminant validity, all women were administered tests from the two general categories of behavior sample and self-report. The CWAS was one of the behavioral tests. In addition, a number of other tests were used. The CWAS was to be compared with other self-report measures of assertion and, in addition, some self-report and behavioral measures of anxiety.

It was found, via Pearson product-moment correlations and a calculation of $\underline{t}$ tests for correlated means (which measure group changes over time), that the CWAS was temporally stable. The results of this study indicated that the CWAS is reliable and sensitive to veridical change.

In terms of interrelatedness of the CWAS and other tests (convergent and discriminant validity), it was found that convergent validity coefficients, which measure the amount of association between different measures, were statistically significant for the assertion measures. This means that the CWAS is associated ($\underline{p}$ < .001) with the College Self-Expression Scale (CSES) (Galassi, Delo, Galassi & Bastien, 1974), the Conflict Resolution Inventory (CRI) (McFall & Lillesand, 1971) and with the Global Rating (this was the authors term for their use of Scale II from the Social Reaction Inventory by Richardson & Tasto, 1976). The study showed that there is convergent validity between these tests of assertion.

Discriminant validity was based upon three progressively more stringent criteria: (a) whether monotrait-heteromethod correlations are higher than heterotrait-heteromethod correlations; (b) whether mono trait-heteromethod correlations are higher than heterotrait-monomethod

correlations and (c) whether the same pattern of relations is shown in all heterotrait blocks independent of method. It was found that only the CWAS test met the first and second of these criteria and that none of them met the third and most stringent of them. Thus, the authors did not discover sufficient evidence of discriminant validity. In sum, Kern and MacDonald provide us with the following information of importance to the use of role-play as an assessment technique: 1) the CWAS role-play method is reliable over time and 2) the CWAS test of assertion correlates with other measures of assertion.

Burkhardt, Green and Harrison (1979) examined the predictive validity and construct equivalence of self-report, behavioral role-play and in vivo assessment of assertive behavior. The College Self Expression Scale (CSES) (Galassi, Delo, Galassi & Bastien, 1974), the Rathus Assertiveness Scale (RAS) (Rathus, 1973) and the Endler S-R Inventory of General Trait Anxiousness (SR-GTA) (Endler, Okada, 1974) were administered to all subjects. The CSES items were separated into groups that represented positive and negative assertiveness. Each subject was interrupted during these paper and pencil self-reports and taken to another room to complete a role-playing task. The role-play required subjects to respond to audiotaped stimulus situations as though they were in the actual situation. The variable used to rate the role-plays were nonverbal behaviors (the latency and duration of response), positive (praise, appreciation and spontaneous behavior) and negative (compliance, request for new behavior and confrontation) content measures and an overall assertiveness measure. The in vivo measure consisted of a phone call by a confederate who was ostensibly soliciting help in preparing for a test in a shared psychology course. The phone call was made approximately two weeks after subjects participation in the study.

The authors found that assertiveness measures are task dependent in that there was more overlap within task than between tasks. The was also a moderate degree of correspondence between self-report (particularly the CSES) and role-playing measures, although this was true only for negative assertion. There was a lack of correspondence between the behavioral measures of assertiveness and the anxiety measures although anxiety was related to the self-report of assertiveness. The authors suggest that differences in levels of anxiety within the laboratory and real life setting may account for the lack of correspondence between these two measure of assertiveness.

In examining positive and negative assertion, the authors found a lack of correspondence between self-report measures of positive and negative assertiveness which was replicated in the small correlations between overall positive and negative assertiveness scores of the role-playing task. Burkhardt, Green and Harrison suggest that positive and negative assertiveness might be seen as unrelated behavioral dispositions that share little of the same behavioral topography. The authors reported that there were no significant correlations between the self-report or role-playing measures and the in vivo measures of assertiveness. Burkhardt, Green and Harrison hypothesized that the telephone in vivo measure lacks adequate content validity as an overall method of assertivenss.

Bordewick nad Bornstein (1980) report an experiment in which they tested the discriminability of aggressive, assertive and non-assertive responses in subjects assessed as high, moderate or low in assertion by the Rathus Assertiveness Schedule (RAS) (Rathus, 1973) and an abbreviated form of a behavioral role-playing assertion test. Seventy-five subjects, twenty-five from each RAS assertiveness level, were given the role-play

assertion test which was scored across six behavioral components and one overall measure of assertiveness (duration of reply, loudness, affect or firmness, latency, compliance, request for change in behavior and overall assertiveness). The assertion discrimination task consisted of 15 written vignettes and replies. These were administered to the subjects. The replies had been previously determined by expert judges to be aggressive, assertive, or non-assertive in nature. On each task the subject rated replies on a nine point Likert type scale in the areas of response classification, likelihood, comfort and valence. The results indicated that subjects scores on the RAS were not strongly related to role-played responses. Aggressive, assertive and non-assertive replies were clearly distinguished among all subjects and individuals differentiated on the basis of assertiveness reached wide agreement regarding response classification. High and moderate assertive individuals demonstrated greater likelihood and comfort in using aggressive and assertive replies than did low assertive subjects. The authors found that assertion levels formed by the RAS exhibited considerable distinctiveness on the likelihood and comfort of the assertion discrimination task, thereby raising questions about the internal and external validity of role-playing assessment instruments.

Galassi and Galassi (1976) investigated the effects of role-playing variations on the assessment of assertive behavior with 48 low assertive college students. The low assertive subjects were randomly chosen from a group who scored low on the College Self Expression Scale (CSES) (Galassi, Delo, Galassi & Bastien, 1974) which is a self-report measure of assertiveness. The subjects were than randomly assigned to 4 of 12 each (6M, 6F) groups--a taped single stimulus group, a taped multiple

stimulus group, a live single stimulus group and live multiple stimulus
group. The subjects in each group role-played six situations in response
to live or taped stimulus situations. Subjects in the multiple stimulus
group were required to respond to five or six stimulus statements per
situation. Subjects'responses were assessed on the basis of assertive
content, response length and anxiety level.

The authors found that mode of stimulus (taped v. live) signifi-
cantly influenced the level of subjective anxiety. Furthermore, the
number of required subject responses significantly affected response
length. Role-playing variations did not significantly affect the content
of responses. Galassi and Galassi mention three implications of their
findings: 1) the kinds of role-playing procedures used for screening
subjects in assertive training research may influence which subjects are
included as assertive or not assertive; 2) if there are assessment
effects following training then the results from studies using different
assessment procedures would not be comparable and 3) nonverbal components
of assertiveness could not be measured by the use of a taped presentation.

Westefeld, Galassi and Galassi (1980) examined the effects of
instructional variations during the role-playing assessment on the
assertive behavior of low assertive college students. Eighty students
(32 males, 48 females) who were assessed as low assertive using the
Conflict Resolution Inventory (CRI) (McFall & Lillesand, 1971) were
randomly assigned to eight assessment groups that differed on two levels
of three factors in a 2X2X2 factorial design. The three factors were:
knowledge of the assertive dimension under study (situational knowledge),
knowledge of the criteria for effective behavior, and knowledge of
instructions concerning how to act (normally v. effectively). The

subjects were assessed for assertion and non-assertion on the basis of
scores from the CRI as well as measures of subject performance during
seven role-playing situations. The role-playing situations were assessed
in terms of content, response latency, response duration, Subjective
Unit of Disturbance (SUDS) (Wolpe & Lazarus, 1966) scale scores and
pulse rate.

The authors found that subjects having knowledge of the situations
taped did not show more assertive content in the role-plays. This was
contrary to their directional hypothesis. Results indicated that situa-
tional knowledge affected pulse rate and self-report anxiety (SUDS)--
subjects who had prior knowledge of the situations had lower pulse rates
and higher SUDS scores. Subjects who had been told of the criteria for
effective behavior had significantly shorter response latencies and
significantly lower response durations than those who were not told this
information.

Subjects told to act effectively had higher assertion scores and
lower non-assertion scores on the CRI posttest than those subjects who
were told to act normally. In general, providing subjects with explicit
information or instructions to behave assertively or effectively, pro-
viding criteria for effective behavior and indicating the types of
role-playing situations during screening or pretesting, resulted in more
assertive behavior than instructions to act normally and instructions
which provided no information about situations and effective behavior.
The authors believe that the implications of these findings is that in
comparing results across assertion training programs one should take
into account the assessment instructions used as instructional variations
do affect performance.

Eisler, Hersen, Miller and Blanchard (1975) investigated, via role-playing, the behavioral components of negative and positive assertion in situations varying in social-interpersonal context. They used 60 male hospitalized psychiatric patients. An additional objective was to delineate behaviors that would differentiate high and low assertive patients in varied contexts of assertion. The Wolpe-Lazarus self-report inventory of assertiveness (Wolpe & Lazarus, 1966) was administered to all subjects. Thirty-two situations containing two levels of three situational variables (positive and negative, male and female, familiar and unfamiliar) were then presented via role-playing in a room containing a T.V. camera and an intercom. Each patient was escorted to the room and seated next to either the male or female interpersonal partner (male or female research assistant) who delivered a predetermined prompt to the patient following directions for each scene from the control room. Patients responses were categorized in terms of nonverbal behaviors (duration of eye contact, smiles, duration of reply, latency of response, loudness of speech, appropriate affect and ratio of speech disturbances to duration of speech); positive content (compliance and request for new behavior) and overall assertiveness.

The authors found that in role-playing situations an individual's behavior was functionally related to the three situational variables. From this they concluded that in situations requiring assertive expression an individual's response is functionally related to the social context of the interaction. Specific results that showed the social context influence were mentioned above in terms of male/female, familiarity and positive/negative. Also, the results demonstrated significant differences in terms of the nonverbal behaviors when the social context

in terms of a positive or negative assertion was varied. More specif-
ically, greater assertion was evidenced toward females in negative
situations than toward men and subjects were more likely to offer praise
and appreciation to female than male partners in positive situations.
Additionally, males elicited different responses than females on all
measures of speech content. Subjects tended to obtain higher overall
assertiveness ratings when their expressions were directed toward un-
familiar persons than compared to familiar persons. Subjects tended to
obtain higher ratings on overall assertiveness when expressing positive
rather than negative feelings.

Eisler, Hersen, Miller and Blanchard believe that their results
support a stimulus specific theory of assertiveness. That is, an indi-
vidual who is assertive in one interpersonal context may not be assertive
in a different interpersonal environment. The authors mention that the
generality of the specific effects of context on assertive behavior
found is limited to the characteristics of the present population and
assertive situations sampled. They contend that it is not likely that
therapists can train clients to be more assertive. Rather, clinicians
should identify classes of interpersonal situations in which deficits
can be identified.

Warren and Gilner (1978) investigated the use of the Behavioral
Test of Tenderness Expression (BTTE) (Warren & Gilner, 1978) to assess
the expression of positive assertive behaviors in 41 dating or married
couples. Each person listened to the BTTE on tape while the other
member of the couple was given the Love Scale (Swenson & Gilner, 1968)
and the Crowne-Marlowe Social Desirability Scale (Crowne & Marlowe,
1960). After both members had completed these three tasks, they listened

separately to tapes of their partners role-play test and rated his or her response on the Partners Rating Form (PRF) (Warren & Gilner, 1978). Subjects were also given a self-report assessment which measured degree and ease of expression of feelings. The Love Scale was used to discriminate what types of love behaviors were measured by the BTTE. The self-report, PRF and Social Desirability Scale were used as measures of concurrent validity.

Warren and Gilner found that the BTTE had high inter-rater reliability. The results showed that the BTTE had high concurrent validity as measured by the partner ratings of taped behaviors. This indicated that the test behaviors were typical of usual home behaviors in similar situations. Social desirability was not related to any of the measures used. Self-report measures yielded moderate correlations with the role-playing test. Over the entire BTTE females scored significantly higher than males. The Love Scale and BTTE did not correlate highly for women, however, the Unexpressed Feelings Factor (UF) on the Love Scale was a significant predictor of the BTTE for males.

Higgins, Alonso and Pendleton (1979) report an experiment in which they attempted to assess the validity of role-play assessments of assertiveness of 16 males and 36 female college students who were equally divided into high and low assertive groups on the basis of assertiveness scores on the Conflict Resolution Inventory (McFall & Lillesand, 1971). Subjects were falsely informed that they would participate with a second subject who had yet to arrive. During this time the College Self Expression Scale (CSES) (Galassi, Delo, Galassi & Bastien, 1974) and a bogus personality inventory were administered to them to reinforce the subjects' belief that they were fulfilling their obligation and would receive

research credit. The experimenter (male) returned later and said that the second subject had still not arrived. In the role-play condition subjects were told that the second (female) experimenter would be in to speak to them but that what she said would not be real. They were instructed to respond as if what the experimenter had said were true. In the actual behavior condition subjects were told only that the second experimenter would be in to see them. The first experimenter left and the second experimenter came in and told the subjects that they would not be given research credit because the bogus second subject had failed to appear. The experimenter then initiated a series of six hierarchically arranged requests in descending order of unreasonableness, supposedly in an effort to reschedule the subjects for another appointment. This was done until the subjects acquiesced once or refused six times. Subjects were assessed on the basis of the results of the request series, a behavioral checklist of behaviors (assertive, aggressive, passive/compliant) performed in the assessment situation and on an overall assertiveness measure. The authors found that actual behavior subjects were significantly more passive/compliant than role-play subjects and male subjects made more passive/compliant responses than females. In the request series, role-play subjects refused more quickly than actual behavior subjects and were reated as more assertive on the overall assertiveness measure.

The authors believe their results suggest that caution should be taken in making inferences about subjects level of invivo assertiveness until the effects of role-play assessments are clarified. The finding that male subjects were more passive/complaint than females in responding to the female experimenter in both conditions parallels previous findings

that subjects are more assertive with same sex than with opposite sex protagonists. The authors believe that role play assessments may be particularly useful in investigating the type and relative magnitude of effects of situational variables on subject assertiveness.

Caldwell (1977) reported a study comparing Behavioral Role Playing Tests (cited in Caldwell, 1977) with taped recorded situations (TRS) (cited in Caldwell, 1977) in the assessment of 32 low assertive males and females. Subjects were selected from a group administered the College Self Expression Scale (CSES) (Galassi, Delo, Galassi & Bastien, 1974). Subjects received both measures with scenes involving three types of content (refusal, request or commendatory assertive responses). Half of the scenes required male and half female role players. Subjects were also given a self-report anxiety measure and a final assessment which was a written paper pencil test of assertiveness using similar scenes. Subjects responses were evaluated on four behavioral components of assertion: response latency, voice inflection, content, and overall assertiveness.

The author found a significant main effect for type of assessment technique on response latency and overall assertiveness. The TRS tended to elicit more assertive behavior than the BRPT. Significant differences were obtained for the sex of the role player and the type of content. Female commendatory scenes elicited the highest assertive responses and female request scenes the lowest for all participants. Analysis of anxiety scores indicated that females had higher anxiety levels than males and all participants showed higher anxiety levels following BRPT assessment. Caldwell stated that the results of this study support the contention that the type of assessment technique is of major importance in determining social skills deficits.

Hester (1978) examined the behavioral validity, using live role-play situations, of the four categories of assertiveness found on the Gambrill-Richey Assertion Inventory (Gambrill & Richey, 1975). The four categories are: assertive, unassertive, doesn't care, and anxious performer. The 48 subjects, who maximally represented, in terms of discomfort and response probability, each category on the assertion inventory were selected from a group of 350 students. All subjects were exposed to six role-play situations, each requiring an assertive response from the subject. Experimental scenes were chosen on the basis of those specific items of the Assertion Inventory which had empirically demonstrated an ability to differentiate among groups. The actual stimulus statements were presented by a live interpersonal partner and subjects responses were tape recorded. Subjects were rated on assertive content, affect, requests for new behavior, anxiety indices and subjective units of disturbance. The results indiated that significance was not obtained on any of the directional hypotheses and no group or sex differences on any of the behavioral measures were found. Hester discussed the results as reflecting upon the clinical usefulness of the Assertion Inventory and upon self-report measures in general.

Alexander (1978) reported a similar study attempting to validate behaviorally through role-play the Gambrill-Richey Assertion Inventory (Grambrill & Richey, 1975). Sample subjects were chosen on the basis of Inventory discomfort and response probability scores which were maximally representative of the four categories of the inventory (assertive, unassertive, doesn't care, and anxious performer). Subjects were individually exposed to six videotaped role-play situations requiring assertive responses. Scene content was determined on the basis of a factor analysis

of the Assertion Inventory with each scene representing a separate factor. Half of the scenes involved a male interpersonal partner and half involved a female interpersonal partner. Subjects responses were taped and rated on four dependent measures: assertive content, affect, request for new behavior, and anxiety indices. The subjects were required to record a Subjective Unit of Disturbance Score (SUDS) (Wolpe & Lazarus, 1966) immediately after responding to each scene.

Alexander reported results that revealed significant differences among the groups on four of the dependent measures (affect, assertive content, request for new behavior, and subjective unit of disturbance). Group differences showed that subjects identified by the Assertion Inventory as being assertive, unassertive, doesn't care or anxious performers were similarily identified by the role-play test on these four measures. The author stated that these findings provide support for the validity of the Gambrill-Richey Assertiveness Inventory as a measure of subject assertiveness.

McFall and Marston (1970) conducted an experiment in which a behavioral rehearsal treatment procedure was used in training 42 subjects to be more assertive. Subjects were recruited from psychology classes by announcing the experiment and allowing students to either phone or register on a sign up sheet. This recruitment procedure made the solicitation procedure less obtrusive and subjects were told that they would not be required to commit themselves immediately. This procedure was used in order to insure a non-assertive group similar to those self-referred in a clinical situation. Subjects were assessed with the use of 16 tape recorded stimulus situations requiring assertive responses. Subjects were told to respond to each situation as if they were actually

in it and that their responses were being tape recorded. The subjects
also rated each situation as to how anxious they would feel if actually
in the situation and how satisfied they would have been with the responses
they had given. Subjects were given a post behavioral test using the
same 16 situations. Subjects were rated on response latency, the asser-
tiveness of the response, and posttest pulse rate. Two weeks following
the post treatment assessment, all the subjects received an in vivo
follow up test which consisted of a phone call by a confederate posing
as a magazine salesman who attempted to sell a subscription. Each call
was rated on resistance to the sale, verbal activity level and social
skill.

Pre and post responses on the behavioral role plays were compared
to determine which was more assertive and as an indicator of improvement
due to treatment. Combined experimental (behavioral rehearsal) groups
showed significantly greater improvement than the combined control
groups. There were no significant differences on post treatment assess-
ment between groups on self-reported satisfaction responses. However,
the combined experimental groups reported a greater reduction in anxiety
and there was a greater decrease in pulse rate than in the combined
control groups. The pattern of results for the in vivo task coincided
with the patterns on the behavioral role-play post assessment. That is,
experimental subjects evidenced a transfer of treatment effects to the
'real life' situation.

Aggression

Rotter and Wickens (1947) attempted to discover the conditions
under which reliable judgments of social aggression can be made using

observations of role-playing situations as an assessment technique. Forty-eight female college students were rated by fourteen judges (divided into teams) on four different roles in two equivalent situations. Each student played two roles and the effect of roles, sequence of roles and patterns in the roles were studied.

Results of the study indicated that the role itself could be a heavy determiner of rating. A social climate was hypothesized to exist between two partners which affected a similar situation in which the same subjects played reverse roles. Rotter and Wickens believe that the selection of specific roles, the specificity of criteria for rating and the partners in a reversed role playing situation may all affect that ratings of trained judges. When these variables are controlled they believe it is possible to obtain consistent ratings in a single situation and moderately reliable ratings from one situation to another.

Leibowitz (1968) conducted an experiment in which the Buss Aggression Machine (BAM) (Buss, 1961), role-playing and the Buss-Durkee Hostility Inventory (BD) (Buss & Durkee, 1957) were compared as assessments of aggression. Leibowitz hypothesized that the three measures would lie along a continuum of behavioral directness with which the aggressive response is made. Thirty-eight male college students were given the BD and Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960) four weeks prior to participating in BAM and role-playing tasks. Each subject was paired with a confederate who allegedly received shock on a bogus discrimination learning task. The accomplice actually received no shock but panel lights on the other side of the screen separating subject and accomplice allowed the accomplice to record the level of shock he would have received. At the end of the BAM task, subjects were taken to

another room and given a series of cards containing a brief description of a situation designed to elicit aggressive responses. They were told to respond as if the situation were occurring at that moment and they were actually speaking to the person in the description. Subjects role-plays were taped and rated on a five point scale of aggression by four judges.

The results of the study indicated that the direct behavioral measure (BAM) did not relate to the BD inventory or to role-playing. There were significant positive correlations between role-playing aggression scores and the BD scale. The hypothesized continuum of responses was not supported by the data. The Marlowe-Crowne yielded significant negative correlations with all measures of aggression as predicted. The author suggests that aggression should be viewed as falling along two dimensions, physical and verbal, and that further research should examine aggression as two modes of expression.

Kirchner, Kennedy and Draguns (1979) report an experiment in which they attempted to assess assertive and aggressive behavior by role-playing and self-report measures in groups of convicted offenders and demographically similar non-offenders. The offender sample consisted of 32 male inmates (16 white and 16 black) from a medium security state correctional institution. The non-offenders (29 males - 14 white and 15 black) were participants in a publicly funded vocational retraining program. These subjects were similar to the offenders in terms of age, IQ, socio-economic background and history of inadequate and irregular employment. Each subject completed three self-report measures; the Adult Self Expression Scale (Gay, Hollandsworth & Galassi, 1975), the Buss-Durkee Hostility Inventory (Buss & Durkee, 1957) and the Social Avoidance and Distress

Scale (Watson & Friend, 1969). In the role-playing task, subjects responded to eight interpersonal conflict situations. Two situations were designed to elicit positive assertive behaviors and one began as a positive situation and changed into one of conflict. Each conflict situation involved a sequence of predetermined lines delivered by a male or female stimulus person with each line being somewhat more provocative than the previous one. Five of the conflict situations involved a familiar person and the other three concerned strangers. Subjects were read the instructions for the role-play and were told to respond as naturally as possible. Subjects were videotaped and all wore white overalls. Raters were told that all participants were prisoners. Subjects responses were rated for aggression, assertion, degree of non-compliance with the demands of the stimulus person, latency of response, duration of response and percentage of maintained eye contact. A count was made of speech dysfluencies, number of attack statements and number of responses made by the participant.

The results of the study revealed a highly significant difference between offenders and non-offenders on the role-playing measures of aggression. Contrary to the authors prediction, there was no significant univariate difference on rated assertion between offenders and nonoffenders. However, an analysis of covariance performed on ratings of assertion (with aggression as the covariate) revealed a significant positive relationship between rated assertion and aggression. When this variability was removed, adjusted means for offenders and non-offenders indicated greater rated assertion among non-offenders. Behavioral and self-report measures of assertiveness were significantly related for only one (the white non-offender) group. A significant relationship

between behavioral and self-report measures of aggression occurred only in a single (black offender) group.

Kirchner, Kennedy and Draguns (1979) believe that the finding with the most potential clinical relevance is that offenders may assert their rights, feelings, wants and goals in conflict situations in a more aggressive manner than non-offenders. The authors stated that the relationships among self-report and behavioral measures of assertiveness may be specific not only to the measures employed but also to the populations being assessed. They suggest that assertiveness training provided for offenders should include and emphasis on decreasing the aggression which accompanies their assertion.

Social Skills

Perri and Richards (1979) employed a research strategy using the behavioral analytic assessment model to create a behavioral role-playing test of heterosocial skills. A situation analysis was carried out to collect a sample of specific heterosocial situations with moderate levels of difficulty and high probability of occurrence. This was done by having undergraduate students complete behavioral diaries about their heterosocial interactions and interviewing ten campus counselors. An independent sample of college males rated this pool of situations for probability of occurrence and level of difficulty. A sample of potential responses to the remaining 60 items was collected using another sample of males and having them respond in each situation. Five females and five males were selected (based on dating patterns and comfortability in social situations) to rate the effectiveness of the potential responses to the situations. The situations that showed the widest disagreement among judges on what constituted an effective response were eliminated.

The remaining 36 situations were used in an assessment format which entailed audiotaped presentations of these situations with a role-playing procedure for response expression. Two contrasted groups, 11 low frequency daters (males) who responded to newspaper ads announcing a treatment program for dating skills and 11 males who dated regularly and rated themselves as competent and comfortable in heterosocial situations, were used as subjects for the assessment procedure. These subjects were administered the Social Avoidance and Distress Scale, the Fear of Negative Evaluations Scale, the Taylor Manifest Anxiety Scale and the Rathus Assertiveness Schedule (cited in Perri & Richards, 1979). Subjects were administered the 36 audiotaped situations and the subjects responses were taped and scored by a pair of independent raters. The subjects verbal replies were compared with sample responses and the effectiveness of the response on a five point basis was scored. The five points being superior (5), high average (4), average (3), low average (2), and inferior (1).

After analysis for inter item consistency, it was found that 14 items did not correlate significantly with the total scale score and these were deleted and the remaining 22 item instrument was named the Heterosocial Adequacy Test (HAT) (Perri & Richards, 1979). Inter and intra-rater reliability coefficients of .85 and .93, respectively, were obtained on the 22 HAT items. Results indicated that the overall test scores of the 'adequate' subjects were significantly better than the 'inadequate' subjects. The discriminant analysis revealed that 100% of the subjects were classified correctly regarding their respective known groups. The HAT scores correlated significantly with the subjects self ratings of heterosocial skills and their scores on the Social Avoidance and Distress Scale. The HAT did not correlate significantly with the

self-report anxiety measures. Perri and Richards feel that additional psychometric evaluations of the HAT appear warranted, particularly the stability of test responses over time, the use of alternate test forms and sensitivity of the HAT as a dependent variable.

Palumbo (1976) investigated the social skills in psychopaths via role-playing. Social skill was defined as the ability to emit and receive positive reinforcement and the ability to emit appropriate behavior in the roles of task and social emotive leaders. Also of interest in this study was the effect of tangible incentives upon the social skills of the group. Seventy-two inmates from a correctional institute were divided into psychopath and non-psychopath groups on the basis of MMPI profiles and Taylor Manifest Anxiety Scale (cited in Palumbo, 1976) scores. Twenty-four three man groups (one psychopath, two non-psychopaths) role-played task and social emotive leaders under tangible and non-tangible incentive conditions. Groups discussed value statements in a warm up and role-play session. The sessions were recorded and scored according to Bales Interaction Process Analysis (cited in Palumbo, 1976). Each subject rated himself and the two other members on task and social emotive leader functions.

Results indicated that none of the hypotheses were supported. Social skills increased equally for both groups from nontangible to tangible incentives, indicating that psychopaths are not more or less socially skilled than others. Psychopaths were as well liked as non-psychopaths and they expressed positive feelings to others to the same degree an nonpsychopaths. A post hoc analysis revealed that psychopaths were more persuasive, emitted more task oriented behavior and were rated higher on task oriented dimensions than the control groups. The differ-

ences occurred when psychopaths held a position on a value statement opposite to that of the other group members. The author suggested that further research should investigate the conditions related to the significant post hoc findings.

Rosenthal (1978) reported a study in which the social skills of deliquent and nondeliquent adolescent girls were compared using a behavioral role-playing inventory. The Problem Inventory For Adolescent Girls (Rosenthal, 1978) was developed using a behavior analytic model for assessing competence in social situations. The rationale behind the research was a hypothesized relationship between deliquent behavior in girls and deficits in social skills. That is, it was hypothesized that some girls may behave maladaptively because they lack the social skills to perform more competently. Competent behavior was defined as the ability to express oneself in a clear and polite way, to share concern for the feelings of others and to defer to adult authority. Incompetent behavior was defined as verbal and physical aggressiveness, acting against the expressed wishes of adults, walking away from problems and committing illegal actions. Groups of deliquent and nondeliquent girls were compared on their inventory performance. Two measures relating to family background--intact v. nonintact families and girls with a history of running versus those without--were compared with inventory performance on those items involving interactions with mothers and fathers.

The results revealed that the test as a whole discriminated between groups of delinquent and nondelinquent adolescent girls, as did 40 of the 52 individual items. Girls from nonintact homes and those with a record of running away from home performed less competently on items involving interactions with parents. The author found that the items on

the inventory were generally independent of one another and 85% of the subjects were assigned to their appropriate delinquent or nondelinquent groups. The author suggests that there is a need to develop and evaluate social skills training programs for delinquent girls.

Bellack, Hersen and Turner (1978) reported two experiments that were conducted to examine the validity of role-play tests of social skills. The first experiment examined the relationship between specific response components on the Behavioral Assertiveness Test-Revised (BAT-R) (Eisler, Hersen, Miller & Blanchard, 1975) and mental health experts ratings of overall response effectiveness and quality. Five female psychiatric patients were administered the BAT-R. Responses required positive and negative assertion and were rated by two sets of raters-- three research assistants and ten mental health experts. Responses were rated for voice volume, eye contact, speech duration, response latency, intonation, overall assertiveness, smiles, expression of praise and appreciation, requests for new behaviors, compliance and the use of physical gestures. Each mental health expert independently rated each response on effectiveness of the response and the quality of the response (how appropriate was the response independently of its effectiveness). Results showed that overall assertiveness ratings on the negative scenes were highly correlated with the two mental health ratings of effective- ness and quality of response. Correlations on the positive scenes were considerably lower but were significant. A multiple regression analysis was used to examine the interactive contributions of the response elements and, again, it was found that the response components on the positive scenes were less predictive of effectiveness and quality of ratings than were response components on the negative scenes. The authors felt that

the results of this experiment lend support for the utility of measuring component elements in evaluating interpersonal skills. They point out that the specific components which make a major contribution to the efficacy of responses will vary according to the types of response (positive v. negative, male v. female). Also, the contributions of components within a response class will vary according to the specific definition of response adequacy or social skill (effectiveness v. quality).

The second experiment examined the correspondence between responses on the BAT-R and two more natural situations (interviews and treatment groups). Ten male and ten female psychiatric patients were observed in three settings: the BAT-R, staged interviews, and a treatment group. The subjects responses on the BAT-R were videotaped and rated as they were in experiment one (described previously). Each subject was individually interviewed on two occasions using open ended questions and these were rated for response latency, speech duration, intonations, and voice volume. Subjects were also observed in two sessions on theme centered and process groups. Initiation of conversation, eye contact while speaking, and talking in response to others were rated on an occurrence/ nonoccurrence basis as well.

The results indicated that BAT-R responses were not significantly correlated with a parallel response in the interview or treatment group situations. The interview and treatment group situations were found to correlate with each other on two measures. The authors mentioned that the BAT-R differed from the structured interview and treatment group situations in content of focus (i.e., assertion) and in terms of form (i.e., role-play v. non role-play). They suggest that any lack of correspondence could be due to either difference. However, the inter-

view and treatment groups also differed substantially in content and yet there was a high relationship between several of the responses in these two situations. The authors believe that the equivocal results between the two experiments could be due to the role-play format. Role-play subjects might not comply with instructions and respond as if the interaction were real. They also mention that chronic psychiatric patients might have special difficulty with the role-play format, making it invalid for them but not for other groups.

Bellack, Hersen and Lamparski (1979) reported a similar study conducted to examine the validity of a role-play test for assessing social skills. Thirty-nine males and thirty-nine females, who fell along a continuum of low to high self-reported dating, completed several self-report inventories and a role-play test of social skill. The role-play test consisted of 20 items in which the situation was described over an intercom and an opposite sex experimental assistant delivered a prompt line to the subject sitting across from him/her. The role-play test consisted of enactments of 20 items developed for heterosocial skill inventories. The items could be translated into role-play scenarios with the addition of a prompt line. The subject was told to respond to the assistant as if the situation were actually taking place. The responses were then videotaped. The subject then reported for a second session and was told that there would be a delay. A few moments later a confederate (who had been trained to be appropriately responsive to any subject comments while limiting their own comments) was ushered in and told the same thing in the subjects presence. This deceptive natural-istic interaction was also videotaped. The role-play test was evaluated for eye contact, response duration, response latency, loudness of speech,

voice intonation, smiles, questions, speech disruptions, and overall effectiveness. The naturalistic interaction was rated for self disclosure, open ended questions, smiles, eye contact, and verbalization.

The results indicated that the role-play test was moderately valid for females and minimally valid for males. Response duration was significantly correlated with open ended questions for males but few other significant correlations appeared. Response duration, questions, speech disruptions, and overall effectiveness were each highly correlated with three or four of the five responses on the naturalistic interactions. However, female subjects were not consistent across situations in their performance on these responses and in most cases of both gender groups less than 25% of the variance was specifically accounted for.

The authors believe that major difficulties in the Bellack, Hersen, and Turner (1978) study were rectified in this experiment. Role-play situations were empirically selected (rather than on an ad hoc basis), the interpersonal situation was closely related to the content of the role-play test and the subject population was more representative of the population at large than in the previous study. Bellack, Hersen and Lamparski mention factors that could limit the generality of their findings. First, the subject's behavior could vary according to the characteristics of the role-play procedure. Secondly, more complex behaviors (e.g., use of social reinforcers) might be more consistent across situations than the responses measured in this study. The authors urge that caution be taken in interpreting research based on role-play procedures until a more definitive picture of their validity emerges.

Rhyne, MacDonald, McGrath, Lindquist and Kramer (1974) devised the Role-Played Dating Interaction (RPDI) instrument to measure social

dating skills. The device is comprised of three discrete four minute encounters with trained confederates following contingency programmed scripts. Two RPDI scores are derived by summing scores across situations for a) specific behaviors designated as appropriate for each interaction and b) observable anxiety behaviors encoded on a Timed Behavioral Checklist (cited in Rhyne, MacDonald, McGrath, Lindquist, & Kramer, 1974).

The instrument was implemented on a group of male college students and the results supported the validity of the instrument. This was done with a contrasted group design in which the RPDI was shown to discriminate between two highly similar groups at the .01 level of significanace. RPDI skill score rater reliability and RPDI anxiety coefficients of agreement were both shown to be high.

Anxiety

Wessberg, Mariotto, Conger, Farell and Conger (1979) reported a study in which they examined the validity of role-plays in assessment of heterosocial anxiety. Forty-five college males were placed in low, medium and high frequency dating groups based on their resposnes to the Social Activity Questionnaire (SAQ) in a pretesting situation. All subjects were involved in two role-plays with a waiting period following each role-play during which the subject was administered the Social Anxiety and Skill Index-A (SASI-A) (Farrell, Mariotto, Conger, Curran & Wallander, 1979). The role-play involved carrying on a conversation with a female confederate in which the subject had been told to imagine he was in a social situation with the confederate. The waiting periods were used as naturalistic observation periods and the subject was falsely told that there was a delay and to behave naturally. The confederates

were all trained to provide relatively standardized and moderately long responses in the interactions. Controls were used for confederate familiarity and role-play order. The SASI-A served as one of the self-report measures and required subjects to rate themselves on the amount of skill and anxiety they actually felt. The Ecological Validity Index (EVI) (Wessberg, Mariotto, Conger, Farrell, & Conger, 1979) served as another self-report and was used to assess the subjects perceptions of the situations they were in and their behavior in those situations. Subjects videotaped responses (role-play and naturalistic observation) were rated by judges on global anxiety and social skill using the Social Anxiety and Skill Index-B (SASI-B) (Farrell, Mariotto, Conger, Curran, & Wallander, 1979) which included behavioral indicators (e.g., rigid posture, excessive arm and hand movement, shaky voice, giving compliments, not allowing long silences, smiling) and produced ratings along seven point dimensions. All subjects were retested on the SAQ after participating in the experiment. The EVI was given at this time as well.

Results of the study showed high inter-rater reliability among judges. Ratings were obtained separately for skill and anxiety. A significant difference in anxiety was found between high frequency (HFD) and low frequency (LFD) dating groups with the HFD group being rated as less anxious. The analysis of social skills ratings showed social skill being rated higher in the role-plays than in the waiting periods. Analysis of the self-reports revealed that high frequency daters rated themselves as more skillful than did low frequency daters. A correlational analysis indicated that judges rank ordered subjects quite similarly across role-plays and waiting periods on the appearance of anxiety and skill. Analysis of the EVI ratings indicated

that subjects rated the waiting period interactions as more like real life and their behavior as more representative of everyday social interactions. The authors believe that their study provides a mixed picture of the ecological validity of role-plays. They feel that if absolute levels of heterosocial skill is under investigation, then high demand role-plays elicit a much higher level of exhibited social skill than do the less constrained more naturalistic waiting periods.

Efran and Korn (1969) reported a study in which they attempted to assess social caution (shyness) using role-play, self appraisal and discussion behavior. The study was divided into three experiments. In experiment I, 93 college males were given the Minnesota Inventory of Social Behavior (MISB) (Williamson & Darley, 1937) as a self-report measure of social anxiety. The subjects were then divided into discussion groups and were told to indicate their preference for 16 tasks, one of which they would have to perform. After preferences were indicated, the groups were left (unknowingly) to leaderless discussion and later each member evaluated the other members on participation. Subjects performed their tasks and then filled out an expectancy questionnaire (EQ) requiring subjects to rate themselves on how they felt compared to other college students on certain skills and abilities. The results indicated that fellow discussion member evaluations correlated highly with evaluations done by an experimenter on who was an active participant (AP) and who was socially cautious (SC). There was a significant difference between AP and SC subjects for two tasks which directly involved verbal, social performance. The EQ differentiated AP and SC subjects and these differentiations were significant.

In experiment II the subjects from the previous experiment were used and asked to role-play several different situations using a telephone as a prop. Subjects were rated on their videotaped role-plays on a four point scale, ranging from an 'inhibited, cautious attempt' to an 'uninhibited, enthusiastic attempt'. The results indicated that the role-playing of AP and SC subjects differed. The role-playing task was found to be significantly related to MISB scores and task performance scores of the previous experiment.

In experiment III 79 different male subjects participated in a discussion group and were classified as AP or SC subjects as in experiment I. These subjects then participated in a bogus taste perception task in an attempt to measure conformity to confederate's judgments about taste. The results showed no difference between the AP and SC groups indicating conformity to the confederates responses. The authors believe that the generality of the behavior tapped by this sociometric classification was evidenced by the differences in role-playing scores. However, they suggest that the kind of social caution measured in the discussion situation relates more to the attainment of peer reinforcement than reinforcement from authority figures and thus limits the generality of the AP-SC classificaton.

Rehm and Marston (1968) conducted an experiment in which they attempted to reduce social anxiety through modification of self reinforcement. Twenty-four college males responded to a verbal announcement offering help for anxiety in heterosocial situations. The subjects were given an initial screening interview which was used to determine that the client did indeed feel he had a problem, that uncontrollable environmental restrictions were not the cause and that the client was otherwise

functioning adequately. Subjects were then assessed using a Situation Test (Rehm & Marston, 1968) made up of alternate forms each consisting of 10 items presented orally on tape. The situation (involving a female) was described by a male and then a line of dialogue was read by a female voice to which the subjects were asked to respond. Subjects were asked to respond aloud as if they would in a real life situation and their responses were recorded. Seven different scales were derived from the experimental subjects responses to the Situation Test and subjects rated their own subjective anxiety for each item. The test was also administered to 12 normal subjects to obtain parallel reliability for subjects anxiety ratings. The recorded responses were independently rated on three dimensions: anxiety, adequacy of response and likability. In addition, recorded responses were scored for average number of words per response, average log latency of response and number of anxiety signs. Subjects were administered a Situation Questionnaire (Rehm & Marston, 1968) in which the amount of discomfort was rated by the subject for each situation. The Manifest Anxiety Scale (Taylor, 1953) and the Fear Survey Schedule (Wolpe & Lang, 1963) were administered as self-report measures of anxiety. The experimental therapy group (SR) involved increasing subjects rate and accuracy of positive self reinforcement using a hierarchy of situations which subjects sought out between sessions. One control treatment (NS) used non-directive techniques. The second control (NT) involved instructions to subjects to work on their own and report weekly. Pretesting and posttesting differences scores yielded greatest improvement for the SR group on self-reports of anxiety and overt behavior, verbal output in the Situation Test and generalization to scores on the Manifest Anxiety Scale.

Zuroff and Schwarz (1978) reported an experiment conducted to assess social anxiety via role-playing. The instrument was constructed by obtaining a pool of 48 anxiety arousing situations (representing four general areas of anxiety) and presenting them to a group of college age volunteers who rated them on a comfortability dimension. Situations with low discomfort ratings or ratings which differed for males and females were eliminated leaving two sets of 10 situations for the test. The situations were tape recorded with a description of the situation and a male or female voice reciting a line of dialogue to which the subjects were to respond. Subjects were given instructions emphasizing that they should try to respond just as if the situation were actually happening. The subjects responses were videotaped behind a one way mirror. Each subject was tested before the start of the experiment and again 11 weeks later. Subjects responses were rated with a behavioral checklist of nine anxiety indicators: head held downcast, abrupt head movements, swaying, extraneous arm and hand movements, arms held rigidly, hands restrained, blocking of speech and extraneous comments. Subjects were also rated for a global measure of anxiety using raters subjective judgments based upon observed behaviors. Subjects were given the S-R Inventory of Anxiousness (Endler, Hunt & Rosenstein, 1962) and the Adjective Check List (Zuckerman, 1960) as measures of anxiety several days prior to pretesting. The Marlowe-Crowne Social Desirability Scale was also administered at this time.

Results indicated that the test-retest reliability over an 11 week period was acceptable. Inter-rater reliability for the tests was high for the behavioral checklist, but fairly low for global rating of anxiety. The behavioral checklist was found to be significantly related to the

global measure of anxiety and the two self-report measures were signif-
icantly related to each other. The checklist score and global rating of
anxiety were significantly less correlated with social desirability than
were the self-report measures. The checklist score was weakly but
positively related to self-report anxiety measures.

Zuroff and Schwarz report two limitations which should be considered
in future research: 1) checklist items were primarily nonverbal behaviors
and did not include more common measures of anxiety and 2) the test is
an analogue situation and does not sample behavior in actual social
situations. They also suggest monitoring one or more of the subjects
psychophysiological responses throughout the test.

Fluid-Rigid Boundaries

Johnson and Quinlan (1980) reported a study using an improvisational
role-playing task to differentiate paranoid from nonparanoid schizophrenics.
The authors hypothesized that paranoid subjects would show more rigid
and exaggerated boundaries in their role-playing behavior, while non-
paranoid subjects will show higher frequencies of confused representa-
tions of character, objects and settings within an unstructured role-
playing task. A boundary refers to distinctions between mental repre-
sentations of self and other, between fantasy and reality, among separate
conceptual realms and between the body and the object world. Thirty-one
diagnosed schizophrenic patients from a long term psychiatric hospital
were divided into paranoid, intermediate and nonparanoid groups on the
basis of hospital diagnosis. Each subject was administered the Diagnostic
Role-playing Test (cited in Johnson & Quinlan, 1980) which consisted of
five roles: bum, politician, teacher, grandparent, and lover. The

subjects were asked to act out what a person in the role would do and responses were videotaped. The resulting scenes were analyzed using the Fluid Boundary Scale (cited in Johnson & Quinlan, 1980). This scale consists of the categories of breaking role, fluidity, intrusion, self-reference, loss of distance and fusion. Scenes were also analyzed using the Rigid Boundary Scale (Johnson & Quinlan, 1980) which consists of the categories of perseveration, concrete replacement, enter/leave, boundary, narration and telephone. The scenes were scored for occurrence of these various items. Also two composite scores, Difference (total fluid minus total rigid) and Percent Fluid (total fluid divided by total fluid plus total rigid with the entire operation multiplied by 100), were obtained for the subjects.

The results indicated that paranoid patients scored significantly higher than nonparanoid patients on the Rigid Boundary measure and significantly lower on the Fluid Boundary measure. The intermediate group was consistently lower on both measures. Both Difference and Percent Fluid measures were found to be good discriminators of paranoid and nonparanoid patients. Johnson and Quinlan believe that rigid and fluid boundaries can be consistently and reliably determined with improvisational role-playing and that paranoid and nonparanoid schizophrenics show differences in their representations of characters, objects and settings.

From the review of related literature we can see that the research on role-playing as an assessment technique is varied and contradictory. It is clear that there is a need for more research on the assessment of social anxiety, particularly assessment utilizing a role-play analogue in conjunction with non-self-report behavioral measures. It has been

pointed out by Zuroff and Schwarz (1978) that the literature on the
precise measurement of social anxiety often focuses on specific anxiety
responses rather than responses to a number of situations or overall
proneness to anxiety. The use of role-playing in combination with a
non-self-report behavioral measure is one method for assessing overall
proness to anxiety.

The present study provides additional information concerning the
use of role-playing in assessing anxiety. The difficulties in eliciting
veridical self-reports of anxiety and the restricted use of these self-
reports indicates a need for a behavioral measure of characteristic
social anxiety such as that developed by Zuroff and Schwarz (1978).
Utilizing the Zuroff and Schwarz Situation Test, the present study
provides additional information concerning the validity of role-playing
as a behavioral assessment of anxiety.

Intercorrelations were made between a behavioral checklist score
and a global rating (both based on the role-plays) and two self-report
measures of anxiety (the S-R Inventory of Anxiousness, Endler, Hunt &
Rosenstein, 1962 and the Adjective Check List, Zuckerman, 1960). In
addition, intercorrelations of these measures with the Marlowe-Crowne
Social Desirability Scale (Crowne & Marlowe, 1960) were obtained.

The essential differences between the present study and that of
Zuroff and Schwarz are: (1) the use of different populations (under-
graduates from Emporia State versus undergraduates from University of
Connecticut), (2) the global ratings of the role plays were made by
different raters than those who used the behavioral checklist, and (3)
the exclusion of a posttest in the present study. Based on the test-
retest reliability findings of Zuroff & Schworz (1978), it was determined

that the use of posttest did not significantly increase the amount and quality of the information obtained relative to the amount of time required.

Chapter 2

METHOD

Subjects

Fifty-five undergraduate students (22 male, 33 female) from intro-
ductory psychology courses at Emporia State University volunteered to
participate as subjects. The mean age of the male and female subjects
was 20.27 and 19.18 years, respectively. All subjects were given the
Zuroff and Schwarz Situation Test (Zuroff & Schwarz, 1978), the S-R
Inventory of Anxiousness (Endler, Hunt & Rosenstein, 1962), the
Adjective Check List (Zuckerman, 1960), and the Marlowe-Crowne Social
Desirability Scale (Crowne & Marlowe, 1960).

Apparatus

The Zuroff and Schwarz Situation Test was the role-play instrument
used in the study. It consisted of twenty audiotaped anxiety-producing
situations to which subjects responded by role-playing. The Situation
Test represented four general areas of anxiety-arousing situations:
professor-student, parent-student, opposite-sex peer and same-sex peer
relationships. All four of the areas were equally represented in the
Situation Test. Videotapes of these role-plays were then rated by one
set of raters using a behavioral checklist and by another set of raters
using a global rating scale.

The audiotaped situations consisted of a male voice describing an
anxiety-producing situation and a male or female voice (depending upon
the sex of the subject) reciting a line of dialogue (prompt line) to
which the subject was to respond (role-play). The audiotape included a
tone signalling the beginning of each segment which was followed by the

37

presentation of the situation and a 30-sec. response interval which
allowed the subject to make his or her response. There was a 3-sec.
delay between the end of a response interval and the presentation of a
new situation. Subject's were required to stand while responding to the
audiotaped situations and their movements were videotaped utilizing an
inconspicuously placed camera in the administration room.

The videotapes of the subject's role-play responses were rated
using the Zuroff and Schwarz (1978) checklist of behavioral anxiety
indicators and their global rating scale. The behavioral checklist
consisted of nine relatively specific behaviors: head held downcast,
abrupt head movements, swaying, extraneous arm and hand movements, arms
held rigidly, hands restrained, failure to reply, blocking of speech and
extraneous comments.

The checklist ratings were made by dividing the 30-sec. response
periods for each of the 20 situations in half, yielding 40 15-sec.
intervals for each subject. The two raters using the behavioral check-
list checked off each behavior that occurred in a given 15-sec. interval.
The score for the checklist was simply the total number of behaviors
checked. The rating was done by two graduate students who volunteered
to participate in the experiment and received training in the use of the
checklist. Training consisted of discussion of the definitions of the
nine behavioral indicators and observation of training tapes. Approxi-
mately 15 hours of training were necessary in order to establish inter-
rater reliabilities exceeding 70% for each of the nine items in the
checklist.

A second measure of the subject's anxiety levels consisted of
subjective judgments (global ratings) based on the observed behavior in

the videotapes. Immediately after viewing each videotape, two separate raters independently assigned the subject a score on a six-point scale: (1) not at all anxious, (2) minimally anxious, (3) somewhat anxious, (4) moderately anxious, (5) highly anxious or (6) very anxious. Two graduate students volunteered to do the global rating and no specific training was given to the two raters as they were to attend to the cues which were ordinarily important in their assessment of anxiety in social situations. These global ratings were included to see if there was a degree of correspondence between the behavioral checklist score and what people typically mean by social anxiety.

In addition to the Zuroff and Schwarz Situation Test, subject's were given three other measures: the S-R Inventory of Anxiousness (Endler, Hunt & Rosenstein, 1962) was a self-report inventory of anxiety given to measure characteristic, overall proneness to anxiety. The Adjective Check List (Zuckerman, 1960) was a self-report inventory given to measure current, overall existent anxiety. The Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960) was a self-report given to determine the tendency to respond in a socially desirable fashion.

Procedure

One week before experimentation subjects were administered the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960). Immediately prior to the role-play assessment, subjects were given the S-R Inventory of Anxiousness (Endler, Hunt & Rosenstein, 1962). After completion of the inventory, the subject was taken to a room equipped with an audiotape recorder and an inconspicuously placed video camera. The experimenter and the subject sat near a small table in the room,

facing the camera, while the subject read the instructions. These instructions emphasized that the subject should try to respond "just as you would if the situation were actually happening" and that his/her performance would be rated for "overall psychological adjustment".

The experimenter then played a demonstration tape of three situations to ensure that subjects comprehended the procedure. The subject was then asked to stand up and back away from the table until she/he was about 12 feet from the camera; the experimenter remained seated during the playing of the recorded situations. Taping of the subject's was made automatically by the videotape recorder in the control room. The entire role-play procedure for each subject took approximately 15 min.

One week after the Situation Test was administered the subject filled out the Adjective Check List (Zuckerman, 1960). Subjects were instructed to check the adjectives which described how they felt "this week".

Chapter 3

RESULTS

Interrater Reliability

Interrater reliabilities for the checklist and global ratings of anxiety were computed using Pearson Product-Moment Correlations. The mean interrater reliability for the raters using the behavioral check list was .88. The mean interrater reliability for the two raters using the global rating scale was .49.

Relationships with Other Measures

Intercorrelations among the four measures of anxiety and their correlation with the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960) were computed using Pearson Product-Moment Correlations. Table 1 presents these correlations.

TABLE 1

Intercorrelations of Four Anxiety Measures and Social Desirability

	2	3	4	Social Desirability
1. Behavioral Check List	.18	.02	.002	-.09
2. Global Rating		-.006	.01	-.01
3. Adjective Check List			.46*	.17
4. S-R Inventory of Anxiousness				.22

Scoring of the Marlowe-Crowne scale has been reversed so that a positive correlation indicates a tendency of those high in social desirability to score lower in anxiety.
*p<.05

It can be seen by looking at the table that the behavioral checklist was not significantly related to any of the other measures of anxiety. The behavioral checklist did not correlate with either of the self-report measures of anxiety or with the global rating of anxiety.

The global rating was not significantly correlated with any of the other measures of anxiety. That is, there was not a significant correlation with either of the self-reports or the behavioral checklist.

The correlation between the two self-report measures (Adjective Check List and S-R Inventory of Anxiousness) was significant (r = .46, p < .01). They were not significantly correlated with the global rating or the behavioral checklist measures of anxiety.

The correlations between the Marlowe-Crowne Social Desirability Scale and the four anxiety measure are also given in Table 1. The Social Desirability measure did not correlate signficantly with any of the anxiety measures. Table 2 presents the means and standard deviations of the various measures used in the study.

TABLE 2

Means and Standard Deviations of Four Anxiety Measures

Measure	Mean	Standard Deviation
Behavioral Check List	130.22	22.92
Global Rating	3.02	.97
Adjective Check List	9.16	3.68
S-R Inventory of Anxiousness	393.76	67.47
Marlowe-Crowne SD Scale	16.02	5.71

Chapter 4

DISCUSSION

The present results indicate a high degree of inconsistency with the results reported by Zuroff and Schwarz (1978). The most notable inconsistency is to be found in the lack of correlation between the global and checklist scores, measures within the same response system. These measures were significantly correlated in Zuroff and Schwarz's research. One explanation for this difference is the methodological differences between the two studies. Zuroff and Schwarz's procedure, utilizing the same raters to make both the global and checklist ratings, creates a methodological confound. This confound has been reduced in the present research by providing a separate set of raters to make the global ratings. It seems probable that the significant correlation between the global rating and behavioral checklist score which Zuroff and Schwartz obtained may be due to the fact that their raters utilized the previously learned behavioral checklist criteria when making the global ratings. Given that the global ratings were included in order to provide a comparison between the behavioral checklist score and what people typically mean by social anxiety, the methodological confounding by Zuroff and Schwarz is especially problematic. The present study indicates that, in fact, the behavioral checklist criteria are not significantly similar to what people typically use when making a judgment about social anxiety.

The correlations between the behavioral measures (behavioral checklist and global ratings) and the two self report measures of

anxiety (S-R Inventory and Adjective Check List) is another area worthy

of comparison. The two self report measures of anxiety were utilized by

Zuroff and Schwartz in order to see if there was any correlation between

overall proneness to anxiety and overall existent anxiety (as measured

by the self report instruments) and social anxiety (as measured by the

behavioral instruments). In both studies, there was no significant

correlation between the Adjective Check List measure of anxiety and the

behavioral measures. This indicates that there is no relationship

between overall existent anxiety and social anxiety as defined by these

measures.

The Zuroff and Schwarz study did, however, indicate that there was

a significant correlation between the S-R Inventory and the global

rating. This would seem to indicate that an individual who has an

overall proneness to anxiety will also be anxious in a social situation.

This finding was not replicated in the present research. This incon-

sistency in the two studies could be due to the confounding by Zuroff

and Schwarz mentioned earlier. Given that the global ratings of Zuroff

and Schwarz were obtained in such a way as to make these ratings ques-

tionable, any further correlations of these ratings with other measures

is automatically suspect. Furthermore, correlating the S-R Inventory

with the behavioral measures of social anxiety would appear to

necessarily produce inflated correlations since the S-R Inventory (as

stated by Zuroff and Schwarz) includes statements pertaining to social

anxiety.

A final point worthy of investigation is the difference in the

global rater interrater reliability in the two studies. The present

study indicated a higher interrater reliability for the global raters

than did Zuroff and Schwarz. Again, this difference could be the result of their methodological confounding. That is, they may have obtained lower interrater reliability for their global raters due to these raters having utilized, to various degrees, the behavioral criteria which they had learned for making behavioral checklist ratings. Rather than relying on the criteria for social norms which they would typically use, these raters may have relied on both these norms and the criteria of the behavioral checklist. This may have influenced the degree to which a common set of social norms for social anxiety (as indicated by high interrater reliability on the global ratings) was revealed as one rater may have utilized more or less of the behavioral criteria than the other.

The results of the present study clearly indicate that Zuroff and Schwarz's contention that the behavioral checklist is an adequate measure of what is generally accepted as social anxiety is questionable. Furthermore, the limitations of this instrument which they mention may be much more serious than they are willing to admit. For example, they mention the limitation that the role-play situation is but an analogue to real life experiences and, yet, state that "one can reasonably assume that subjects who are anxious in role-played situations are unlikely to be free from anxiety in their daily lives" (p. 378). Studies by Bellack, Hersen and Turner (1978); Bellack, Hersen and Lamparski (1979); Higgins, Alonso and Pendleton (1979); and Burkhardt, Green and Harrison (1979) make such a statement highly suspect. These studies have shown that, in fact, there is very little relationship between an individuals performance in an analogue situation and an in vivo or real life setting.

The present study also contradicts the statement by Zuroff and Schwarz that the results of their study are similar to that of Lang (1969). Lang found that anxiety measure were not significantly correlated with each other unless they were in the same response system. The significant correlations of the Adjective Check List and the S-R Inventory of Anxiousness do support this hypothesis. However, the low correlation between the behavioral checklist and the global rating in the present research does not lend support to it.

Zuroff and Schwarz recommend that future research should extend the technique to include verbal behavior and physiological responses. While this may be a useful tact, the present study indicates that considerable work has yet to be done on the nonverbal behavior measure as well. The underlying assumption of the study has been that the global rating is an adequate measure of social anxiety and that by correlating the behavioral checklist with this measure, the validity of the behavioral checklist as a measure of social anxiety can be determined. It would appear to this author that there is insufficient evidence to support the use of the global raters' conceptions of social anxiety in both studies.

REFERENCES

REFERENCES

Alexander, D.H. Role-playing validation of the Gambrill-Richey Assertion Inventory. <u>Dissertation Abstracts International</u>, 1978, <u>39</u>(11-B), 5532.

Arkowitz, H., Lichenstein, E., McGovern, K., & Hines, P. The behavioral assessment of social competence in males. <u>Behavior Therapy</u>, 1975, <u>6</u>, 3-13.

Bellack, A.S., Hersen, M., & Lamparski, D. Role-play tests for assessing social skills: Are they valid? Are they useful? <u>Journal of Consulting and Clinical Psychology</u>, 1979, <u>47</u>, 335-342.

Bellack, A.S., Hersen, M., & Turner, S.M. Role-play tests for assessing social skills: Are they valid? <u>Behavior Therapy</u>, 1978, <u>9</u>, 448-461.

Bordewick, M.C., & Bornstein, P.H. Examination of multiple cognitive response dimensions among differentially assertive individuals. <u>Behavior Therapy</u>, 1980, <u>11</u>, 440-448.

Burkhardt, B.R., Green, S.B., & Harrison, W.H. Measurement of assertive behavior: Construct and predictive validity of self-report, role-playing and in vivo measures. <u>Journal of Clinical Psychology</u>, 1979, <u>35</u>, 376-383.

Buss, A. <u>The psychology of aggression</u>. New York: Wiley, 1961.

Buss, A., & Durkee, A. An inventory for assessing different kinds of hostility. <u>Journal of Consulting and Clinical Psychology</u>, 1957, <u>21</u>, 343-348.

Caldwell, A.T. A comparison of assessment modes with low assertive men and women. <u>Dissertation Abstracts International</u>, 1977, <u>38</u>,(11-B), 5560.

Crowne, D.P., & Marlowe, D. A new scale of social desirability
    independent of psychopathology. Journal of Consulting Psychology,
    1960, 24, 349-354.

Efran, J.S., & Korn, P.R. Measurement of social caution: Self
    appraisal, role-playing and discussion behavior. Journal of
    Consulting and Clinical Psychology, 1969, 33, 78-83.

Eisler, R.M., Hersen, M., Miller, P.M., & Blanchard, E.B. Situational
    determinants of assertive behaviors. Journal of Consulting and
    Clinical Psychology, 1975, 43, 330-340.

Endler, N.S., Hunt, J.McV., & Rosenstein, A.J. An s-r inventory of
    anxiousness. Psychological Monographs, 1962, 76(17, Whole No. 536).

Endler, N.W., & Okada, M. An s-r inventory of general trait anxiousness
    (Report No. 1). Toronto, Canada: York University, Department of
    Psychology Reports, 1974.

Farrell, A.D., Mariotto, M.J., Conger, A.J., Curran, J.P., & Wallander,
    J.L. Self ratings and judges ratings of heterosexual social
    anxiety and skill: A generalizability study. Journal of
    Consulting and Clinical Psychology, 1979, 47, 164-175.

Galassi, M.D., & Galassi, J.P. The effects of role-playing variations
    on the assessment of assertive behavior. Behavior Therapy, 1976,
    7, 343-347.

Galassi, J.P., Delo, J.S., Galassi, M.D., & Bastien, S. The college
    self expression scale: A measure of assertiveness. Behavior
    Therapy, 1974, 5, 165-171.

Gambrill, E.D., & Richey, C.A. An assertion inventory for use in
    assessment and research. Behavior Therapy, 1975, 6, 550-561.

Gay, M.L., Hollandsworth, J.G., & Galassi, J.P. An assertiveness
    inventory for adults. Journal of Counseling Psychology, 1975, 22,
    340-344.

Hester, R.F. A role-playing validation of the Gambrill-Richey Assertion Inventory using live interpersonal partners. Dissertation Abstracts International, 1978, 39(9-B), 2150-4581.

Higgins, R.L., Alonso, R.R., & Pendleton, M.G. The validity of role-play assessments of assertiveness. Behavior Therapy, 1979, 10, 655-662.

Johnson, D., & Quinlan, D. Fluid and rigid boundaries of paranoid and nonparanoid schizophrenics on a role-playing task. Journal of Personality Assessment, 1980, 44, 523-531.

Kern, J.M., & MacDonald, M.L. Assessing assertion: An investigation of construct validity and reliability. Journal of Consulting and Clinical Psychology, 1980, 48, 532-534.

Kirchner, F.P., Kennedy, R.E., & Draguns, J.G. Assertion and aggression in adult offenders. Behavior Therapy, 1979, 10, 452-471.

Leibowitz, G. Comparison of self-report and behavioral techniques of assessing aggression. Journal of Consulting and Clinical Psychology, 1968, 32, 21-25.

MacDonald, M.L. Measuring assertion: a model and a method. Behavior Therapy, 1978, 9, 889-899.

McFall, R.M., & Marston, A.R. An experimental investigation of behavioral rehearsal in assertive training. Journal of Abnormal Psychology, 1970, 76, 295-303.

McFall, R.M., & Lillesand, D.B. Behavioral rehearsal with modeling and coaching in assertion training. Journal of Abnormal Psychology, 1971, 77, 313-323.

Palumbo, S.J. An investigation of social skills in psychopaths in terms of role-playing. Dissertation Abstracts International, 1976, 37 (2-B), 981.

Perri, M.G., & Richards, C.S. Assessment of heterosocial skills in
    male college students. Behavior Modification, 1979, 3, 337-354.

Rathus, S.S. A 30 item schedule for assessing assertive behavior.
    Behavior Therapy, 1973, 4, 398-406.

Rehm, L.P., & Marston, A.R. Reduction of social anxiety through
    modification of self reinforcement: an instigation therapy
    technique. Journal of Consulting and Clinical Psychology, 1968, 32,
    565-574.

Rhyne, L.D., MacDonald, M.L., McGrath, R.A., Lindquist, C.U., & Kramer,
    J.A. The role-played dating interactions (rpdi): an instrument
    for the measurement of male social dating skills. Catalog of
    Selected Documents in Psychology, 1974, 4, 42-43.

Richardson, F.C., & Tasto, D.L. Development and factor analysis of
    social anxiety inventory. Behavior Therapy, 1976, 7, 453-467.

Rosenthal, L. A comparison of social skills in deliquent and
    nondeliquent adolescent girls using a behavioral role-playing
    inventory. Dissertation Abstracts International, 1978, 39(9-B),
    4594.

Rotter, J.B., & Wickens, D.D. The inconsistency and generality of
    ratings of "social aggressiveness" made from observation of role-
    playing situations. American Psychologist, 1947, 2, 333.

Taylor, J.A. A personality scale of manifest anxiety. Journal of
    Abnormal and Social Psychology, 1953, 48, 285-290.

Warren, N.J., & Gilner, F.H. Measurement of positive assertive
    behaviors: The behavioral test of tenderness expression. Behavior
    Therapy, 1978, 9, 178-194.

Watson, D., & Friend, R. Measurement of social-evaluation anxiety. Journal of Consulting and Clinical Psychology, 1969, 33, 447-457.

Wessberg, H.W., Mariotto, M.J., Conger, A.J., Farrell, A.D., & Conger, J.C. Ecological validity of role-plays for assessing heterosocial anxiety and social skill of male college students. Journal of Consulting and Clinical Psychology, 1979, 47, 525-535.

Westefeld, J.S., Galassi, J.P., & Galassi, M.D. Effects of role-playing on assertive behavior: a methodological study. Behavior Therapy, 1980, 11, 271-277.

Williamson, E.G., & Darley, J.G. The Minnesota Inventory of Social Behavior. New York: Psychological Corporation, 1937.

Wolpe, J., & Lang, P.J. A fear survey schedule for use in behavior therapy. Behavior Research and Therapy, 1963, 2, 27-30

Wolpe, J., & Lazarus, A.A. Behavior Therapy Techniques: A Guide to the Treatment of Neuroses. New York: Pergamon, 1966.

Zuckerman, M. The development of an affect adjective check list for the measurement of anxiety. Journal of Consulting Psychology, 1960, 5, 457-462.

Zuroff, D.C., & Schwarz, J.C. An instrument for measuring the behavioral dimension of social anxiety. Psychological Reports, 1978, 42, 371-379.