Dow, Michael G.

PEER VALIDATION AND IDIOGRAPHIC ANALYSIS OF SOCIAL SKILL DEFICITS

The Pennsylvania State University PH.D. 1983

The Pennsylvania State University

The Graduate School

Department of Psychology

Peer Validation and Idiographic Analysis of Social Skill Deficits

A Thesis in

Psychology

by

Michael G. Dow

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 1983

Michael G. Dow

Michael G. Dow

We approve the thesis of Michael G. Dow.

Date of Signature:

7/15/83 — W. Edward Craighead

W. Edward Craighead
Professor of Psychology
Chairman of Committee
Thesis Adviser

7/15/83 — Robert M. Stern

Robert M. Stern
Professor of Psychology
Head of the Department of Psychology

7/15/83 — Thomas D. Borkovec

Thomas D. Borkovec
Professor of Psychology

7/15/83 — Robert E. Kennedy

Robert E. Kennedy
Affiliate Assistant Professor
of Psychology

7/15/83 — Linda W. Craighead

Linda Wilcoxon Craighead
Associate Professor of
Counseling and Educational Psychology

ABSTRACT

It is generally considered that few specific behavioral components of social competence have been identified. Prior research in this area is reviewed, and a somewhat more optimistic conclusion is reached. There appear to be nine behavioral codes which have been significantly related to social competence in two or more studies: eye contact, smiles, verbal personal attention, trembling-stammers-fidgets, talking time, hand gestures, pauses or long latencies, compliments, and positive statements about other people or things. Research results have been inconsistent across studies, however, so it is difficult to know whether these behaviors are reliably related to social competence. Moreover, past research has generally ignored the possibility of individual differences among socially inadequate people.

The difficulty of identifying behavioral components of social skill seems paradoxical when compared with the relative ease with which naive observers can reliably distinguish subjects who are high vs. low in self-reported social competence. Thus, the present study employed an assessment methodology which relied on peer judgments to identify specific problematic conversational behaviors. Forty-two subjects, both male and female, who reported high social avoidance and distress, each interacted with three different randomly assigned peers of the opposite sex in dyadic social conversations. The 126

peers completed rating scales of the nine behavioral categories listed above, indicating both the degree and direction of suggested change in their partner's conversational behavior.

Results indicated that each of the nine behaviors was rated as more important for change than a group of comparison behaviors which had no previous research support. In addition, peers reliably evaluated the changes that they would recommend on six of the nine behavioral scales. An analysis of these six measures indicated that peers recommended different changes for different individuals. Three conceptually important subject groups were identified: those with general difficulties across the set of behaviors (11.9%), those with specific differential difficulties (28.6%), and those with no reliably indicated overt social skill problems (59.5%). Thus, this research generally corroborates the importance of many of the previously identified behaviors, although individualized assessment and treatment of social inadequacy seems warranted.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank the members of my doctoral committee for their assistance and support during the development and execution of this project: Dr. W. Edward Craighead, Dr. Thomas D. Borkovec, Dr. Robert E. Kennedy, and Dr. Linda Wilcoxon Craighead. I would especially like to thank my academic adviser, Dr. W. Edward Craighead, who has guided my professional development during these graduate years at Penn State. I would also like to thank Lisa Bailyn, who competently served as a research assistant and contributed numerous hours toward the completion of this project.

CHAPTER I

INTRODUCTION

The Relevance of Social Competence for Clinical Psychology

Within the past 30 years, increasing attention has been focused on the role of interpersonal behavior in the etiology and treatment of psychological problems. Sullivan (1953) was among the first to stress the relevance of interpersonal behavior for clinical psychology, while Zigler and Phillips (1961) are often credited with the thesis that social competence, per se, is of central relevance to clinical problems.

Zigler and Phillips did not investigate the active process of interpersonal behavior, but chose to employ variables such as age, education, intelligence, employment history, marital status, and occupation as summary indices of the construct of social competence. Zigler and Phillips (1961) found that psychiatric patients who were relatively low in social competence averaged a longer period of institutionalization and were more likely to be rehospitalized than patients who were relatively high in social competence.

While the work of Zigler and Phillips was instrumental in emphasizing the relevance of social competence for psychopathology, the variables they used as summary indices of this construct were far removed from the active dynamics of interpersonal behavior. Later

formulations of social competence conceptualized that the execution of interpersonal behavior involved complex motor skills (e.g., Argyle & Kendon, 1967). However, it was not until Lewinsohn (Lewinsohn, 1974; Lewinsohn, Weinstein, & Shaw, 1969) proposed a theory of depression which posited the etiological role of social skill deficits, that specific ties, linking the client's interpersonal behavior with a form of psychopathology, were drawn. Lewinsohn argued that depression is caused from a low rate of response contingent positive reinforcement. An individual's level of social skill was deemed to be of instrumental importance in determining the amount of response contingent reinforcement which would be received from the environment. Indeed, Libet and Lewinsohn (1973) operationally defined social skill as the complex ability to emit behaviors which are reinforced by others and to refrain from emitting behaviors which are punished or extinguished by others.

During the same period of time in which Lewinsohn developed his theory of depression, numerous individuals were vocally questioning the relevance and importance of various target behaviors within the emerging field of behavior therapy (e.g., Cooper, Furst, & Bridger, 1969). Studies on snake phobia drew the heaviest criticism. These concerns led Borkovec, Stone, O'Brien, and Kaloupek (1974) to propose several criteria which might be used to judge the clinical relevance of various target problems. These criteria included: (1) a clinically relevant target behavior should occur at some reasonable frequency among the psychiatric population, (2) the target behavior

should be a source of discomfort or interference in one's daily life, (3) the target behavior should not be influenced by demand or suggestion effects, (4) physiological arousal should occur in anticipation of, and in response to, the feared situation, and (5) there should not be marked habituation of physiological arousal upon exposure to the feared situation.

In contrast to small animal phobias, social anxiety has been shown to meet substantially the above criteria. Borkovec et al. (1974) recommended that social anxiety be considered an appropriate target behavior for clinical research and treatment. From that time period to the present, the related constructs of speech anxiety, social anxiety, minimal dating, communication apprehension, and social skill deficits have received extensive empirical and theoretical attention (cf. Arkowitz, 1981).

Curran (1977) presented the first major review of the area of heterosexual-social anxiety, in which he argued that this megaconstruct might best be viewed as involving three etiological processes: conditioned anxiety, skills deficits, and faulty cognitive-evaluative self-appraisals. This organization has generally been endorsed by other authors and has contributed to the recent practice of referring to this general problem construct as "social inadequacy," instead of "social anxiety" or "heterosexual-social anxiety" (see Arkowitz, 1981; Curran & Wessberg, 1981).

Most of the research and theoretical work in the area of social inadequacy has stressed the role of overt behavioral deficits, although

recent attention appears to be shifting somewhat toward examining cognitive processes (Dow & Craighead, in press). Social skills training has become a widely employed and relatively effective therapeutic technique (Twentyman & Zimering, 1979). Empirical research and conceptual models have stressed the relevance of overt social behavior for such disparate problem constructs as alcoholism (Marlatt & Donovan, 1981), depression (Libet & Lewinsohn, 1973), chronic mental patients (Hersen, 1979), marital distress (Gottman, 1980), and isolated or withdrawn children (Conger & Keane, 1981).

Socially inadequate behavior appears to be fairly prevalent within the psychiatric population. Curran, Miller, Zwick, Monti, and Stout (1980) found that 7.4% of a large group of psychiatric inpatients were judged by a professional staff to be socially inadequate, and Bryant, Trower, Yardley, Urbieta, and Letemendia (1976) found that 16.3% of a large outpatient sample were judged to be socially inadequate. Moreover, the prevalence of inadequate social behavior appears to be fairly uniform among the major psychological disorders (Bryant et al., 1976; Curran et al., 1980; Youngren & Lewinsohn, 1980).

The Search for Behavioral Components of Social Skill

Despite the enthusiasm with which behaviorists adopted "social skill" as a relevant target behavior, there has been a tendency to ignore important definitional issues and/or to assume a high degree of behavioral consistency across situations. With the publication of

several reviews in the late 1970s (Bellack, 1979; Curran, 1977, 1979), behaviorists were put on notice to clarify their unit of analysis and domain of generalization when discussing behavioral assessment and training of social skill. Social skills programs during the 1970s tended to teach behaviors such as eye contact, smiles, questions, and compliments, although there was little research support for the social skill relevance of these behaviors, and skills training was often taught uniformly without adequate consideration of situational relevance.

The realization by behaviorists that they may have inadvertently endorsed trait methodology led to a crisis of confidence, of sorts, which set off a plethora of assessment-oriented research in the late 1970s and early 1980s.

Within the last three or four years, there appears to be a growing recognition of the importance of situational specificity in behavioral assessment (Kazdin, 1979). Individuals may act differently in different situations. Therefore, a newly accepted assessment strategy involves an initial investigation to determine which situations are often problematic for individuals with a particular problem construct. Then, assessment relative to those situations is conducted (see Goldfried & D'Zurilla, 1969). Within the area of social inadequacy, essentially all research which has attempted to identify behavioral components of social skill has involved first conversations with strangers. This social situation does appear to be particularly difficult (Bryant & Trower, 1974), although the

generalizability of research findings to other settings remains largely unclear.

Despite widespread interest and burgeoning research during the past decade, it is generally considered that relatively few behavioral components of social skill have been identified (Conger & Farrell, 1981; Dow, Glaser, & Biglan, 1981). However, this is an extremely difficult area of research to review. One's degree of optimism will vary depending on the subjective weighting of numerous factors. There are at least 18 studies in this area, which vary with respect to: (1) the number and type of behaviors investigated, (2) the use of analogue vs. clinical subjects, (3) the number and type of conversational tasks, (4) the definitions of particular behavioral codes, (5) subject selection criteria, (6) statistical procedures, and (7) the overall type of research methodology employed (e.g., comparisons of known groups, correlational studies, experimental investigations).

There have been at least 13 group comparison studies, where subjects have usually been selected based on self-reported differences in social skill, dating frequency, or social anxiety (viz., Arkowitz, Lichtenstein, McGovern, & Hines, 1975; Barlow, Abel, Blanchard, Bristow, & Young, 1977; Borkovec, Fleischmann, & Caputo, 1973; Borkovec et al., 1974; Cherulnik, Neely, Flanagan, & Zachau, 1978; Glasgow & Arkowitz, 1975; Greenwald, 1977; Jaremko, Myers, Daner, Moore, & Allin, 1982; Martinez-Diaz & Edelstein, 1980; Rehm & Marston, 1968; Trower, 1980; Twentyman & McFall, 1975; Dow,

Biglan, & Glaser, Note 1). There have been at least three correlational analyses which attempted to relate the frequency of certain conversational behaviors to ratings of social skill or interpersonal attraction made by observers or interaction participants (viz., Bayes, 1972; Conger & Farrell, 1981; Kupke, Hobbs, & Cheney, 1979). And finally, at least two studies employed experimental manipulations of specific behaviors and assessed the impact on others (viz., Dow et al., 1981; Kupke, Calhoun, & Hobbs, 1979).

If one takes a broad look at these studies, there are nine subject behaviors which have each been related to social competence in at least two studies: more eye contact, more smiles, more personal attention (e.g., talking about the other and asking questions), less overt or verbal indicators of anxiety such as trembling, stammers, or fidgets, more talking time, more hand gestures, fewer pauses or long latencies, more compliments, and more uppers (positive statements about third persons or things). However, such a broad view ignores the possible Type I error problems which are associated with multiple studies, multiple types of assessment within studies, multiple codes within studies, the use of several related versions of the same basic code in some studies, and the bias of not publishing nonsignificant results.

Table 1 shows the success ratios of the nine codes listed above. The first column lists the proportion of studies where a code was successfully related ($p < .05$) to social competence at least once. This ratio ignores all of the noted potential Type I error problems

Table 1

Potential Social Skills With the Most Research Support

Behavior	Study Success Ratio	Instance Success Ratio
Eye contact	3/7	5/13
Smiles	3/4	4/5
Verbal personal attention	2/2	2/2
Trembling, stammers, fidgets	4/5	4/11
Talking time	8/13	10/22
Hand gestures	2/3	3/4
Pauses or long latencies	3/7	5/12
Compliments	2/2	2/3
Uppers (positive statements)	2/2	2/4

except the use of multiple studies. The second column reports the proportion of coded instances in these studies in which a behavior was related to social competence. This second ratio takes into consideration all of the above noted potential Type I error problems except for publication bias and the coding of multiple behaviors within studies. However, this ratio is also biased in unknown ways by the fact that studies differ in the number of times the same behavior is coded. Thus, this ratio combines independent and non-independent results.

This summary has also ignored, up to this point, the potential advantages and problems associated with particular methodologies. Discussion of particular methodologies will be considered later, but the point remains that despite the tendencies among reviewers to evaluate conservatively the results of this research, there is some evidence for the social skill relevance of each of these codes. At this point in time, multivariate analyses are urgently needed, or perhaps these same behaviors might be revalidated using a different methodology.

The difficulty of identifying relevant behaviors seems paradoxical when compared with the relative ease with which subjects who differ in self-reported social adequacy can be successfully differentiated on global observer ratings of social skill (e.g., Arkowitz et al., 1975; Greenwald, 1977). Curran's (1979) lament seems most appropriate, "Everyone seems to know what good and poor social skills are but no one can define them adequately" (p. 321).

This paradox led Conger, Wallander, Mariotto, and Ward (1980) to investigate which aspects of social interaction were reportedly used by naive observers when making judgments of social skill and anxiety. An extensive list of cues was generated--most of which were similar to the variables being investigated in this area, although conversational style and content were emphasized to a greater extent. While there was no evidence concerning whether these observers were actually influenced by the behaviors they reported, this study stands out as a particularly creative initial attempt to approach the identification of social skills using a different methodology. Indeed, each of the methods which have been used to identify social skill deficits are not without serious shortcomings (Galassi & Galassi, 1979).

The most common assessment procedure utilizes an analogue conversational task, where comparisons are made of known groups who are low vs. high in dating frequency, self-reported social skill, or self-reported social anxiety. Extreme groups may differ in extraneous ways, however, so there is no guarantee that identified behaviors are relevant to social competence. A second method of nominating social skills involves correlating observer ratings of social skill with coded frequencies of specific social behaviors--usually within a normal sample. However, demonstrating such a correlation does not ensure that the relevant behavior is causally related to the skill rating. Other unknown factors might be responsible for variation in both measures. A further concern with both of these methods is whether the types of conversation which have been coded in these studies can be generalized to conversations (even first conversations) with peers

in the subjects' natural environment. Many of these studies utilized trained confederates who were selected to be attractive and socially skilled. Hence, there may be a structural bias whereby the confederates were more similar to the high competent subjects than to the low competent subjects. Any behavioral or observer rating difference between groups could be partially accounted for by differences in degree of similarity with the confederates (cf. Jaremko et al., 1982). The third method of identifying social skills involves manipulating the frequency or topography of specific conversational behaviors and assessing the impact of these manipulations on naive subjects. This methodology represents the essence of clinical treatment research, and as such, it is the most conclusive way, but by far the most difficult, of establishing the clinical relevance of specific social behaviors. Unfortunately, treatment studies almost always combine a number of interventions, whose independent effects, if present, cannot be deciphered.

In summary, overt social competence appears to be a relevant assessment and treatment target for clinical psychology, but research attempting to identify behavioral components of social competence has been hampered by numerous methodological and definitional complexities.

Suggestions for Future Research

Given the imperfection of any methodology and the inconclusiveness of the results generated by the most frequently used methods, it seems appropriate that a peer validation approach, in the spirit of

Conger et al. (1980), be reconsidered and expanded. Peer validation offers the potential advantage of nominating particular social behaviors which are judged to be relevant by the same class of individuals with whom a client would like to interact more favorably. Hence, treatment components derived from this methodology might be especially generalizable to the client's natural environment.

A more subtle concern with previous attempts to identify specific social skills is that there appears to be a uniformity assumption, or "uniformity myth" (Kiesler, 1966), in much of this research. The first point to be considered in this regard is that the behaviors which can improve the social skills of clinical subjects need not be the same behaviors which are associated with social skill among normal subjects or among those individuals high in social skill. It may be that therapeutically increasing a few classes of behavior which are prevalent among high competent subjects will not fit in well, or will otherwise be ineffective, in the configurational context of a client's verbal behavior, nonverbal behavior, and physical appearance. For example, smiling has been related to social competence in three of four studies (viz., Arkowitz et al., 1975; Bayes, 1972; Conger & Farrell, 1981; Trower, 1980); however, research has also shown that social skill is positively associated with physical attractiveness (Greenwald, 1977). Thus, it is possible that newly developed smiles among subjects low in social skill (and less attractive) might not be as effective, on average, as smiles by socially competent (and more attractive) subjects. Thus, it remains unclear whether increasing

smiles would be an effective treatment component among subjects low in social skill. A second example is that compliments have been positively related to social skill in two of two studies (viz., Dow et al., 1981; Dow et al., Note 1). However, given the fairly well-replicated finding that low skilled subjects average less talking time than subjects higher in social skill (e.g., Glasgow & Arkowitz, 1975; Greenwald, 1977), it might seem artificial for the frequency of compliments to increase without a previous increase in the general rate of talking, or without increases in other behaviors which might indicate interest, such as questions and talking about the other person.

In other words, the attempt to identify behaviors which differentiate subjects who are high vs. low in social skill, and the attempt to determine specific behaviors which covary with social skill across a range of normal subjects, are actually indirect attempts to answer a somewhat different question: "What behaviors can improve the social skill of clinical subjects?" It is suggested that clinically related research be designed with this latter question in mind. Before the known groups approach and the correlational approach can be considered relevant to the treatment question, one must assume a certain uniformity of desirable social behavior, regardless of the configurational context in which that behavior is viewed.

Strict uniformity of desired social behavior seems unlikely, although this question has not been addressed by researchers in the area of social skill. Some previous research does appear relevant

to this question, however. Ellsworth and Carlsmith (1968) factorially manipulated the content of an interviewer's speech to be either favorable or unfavorable toward the subject, and they manipulated the interviewer's level of eye contact to be either high or low. Results indicated a significant interaction such that subjects who had received favorable content liked the interviewer more if there had been high eye contact, while subjects who received negative content liked the interviewer more if there had been low eye contact. Thus, the results of this study are consistent with a configurational model of person perception. In other research, Cantor and Mischel (1979a, 1979b) have investigated the tendencies of observers to categorize people into specific personality prototypes. They found, for example, that subjects could more readily recall information about hypothetical persons if the listed personality descriptions formed a consistent prototype. Cantor and Mischel argued that category information can function as an organizational theme which structures the encoding of new information, provides expectations about future behavior, and facilitates the retrieval of past information. Given these findings by Cantor and Mischel, one might speculate that individuals would be more comfortable when interacting with people who are consistent in their self-presentation and behavior. Thus, desired modifications in the behavior of a socially inadequate person might be quite different from the behaviors that have been related to social skill among socially competent subjects. More specifically, some of the social

skills among competent subjects might be too advanced (e.g., compliments) or might otherwise seem incongruous when emitted by socially inadequate individuals.

The second concern related to the uniformity myth is that there may be relevant individual differences among clinical subjects low in social skill. The hallmark of behavior therapy is individualized assessment, where a functional analysis is conducted and a treatment plan is implemented to remediate the particular problems of an individual client (Craighead, Kazdin, & Mahoney, 1981; Kanfer & Saslow, 1965). While assessment research has begun to deal with the thorny question of situational specificity, there appear to be no attempts to include the potential ramifications of individual differences in assessment-oriented research, although social skills treatment research has begun to deal with this concern (e.g., Marzillier & Winter, 1978). At issue is the possibility raised by Cronbach (1957, 1975) that subject characteristics may interact with treatment components. Subsets of socially inadequate individuals may benefit differentially depending on the treatment. Curran's initial classification of social inadequacy into skills deficits, conditioned anxiety, and relatively strict cognitive evaluations was an initial attempt at dealing with this issue. However, within a group of subjects who are judged to have skills deficits (for example), there may be subgroups who would benefit differentially by stressing particular types of conversational behavior.

Given the historical perspective of functional analysis, it seems important to reconsider the role of group assessment studies in behavior therapy. The value of these studies, it seems, is to offer information to the clinical assessor on the base rates of specific deficits within a general problem construct such as social inadequacy. It would be helpful to know what behaviors, on average, could be targeted to improve the social skill of relevant clinical subjects. Then the clinician must decide whether to offer a packaged program employing stock interventions, or to attempt individualized assessment and treatment. Logic and convenience would suggest that clinicians should stick with base rate evidence unless individual assessment and treatment can produce greater success than packaged programs applied routinely (cf. Cronbach, 1957). However, no study has investigated whether clinicians, peers, or other individuals can agree which social behaviors should be modified in particular clients. Nor have studies compared the therapeutic effectiveness of packaged versus individualized programs.

It may be especially important to consider social skills from an idiographic point of view. Some clients may talk too much--others too little, some may exhibit too much eye contact--others not enough. Moreover, the amount of eye contact which would be judged desirable by others may vary depending on one's physical characteristics or other aspects of one's conversational style.

It appears that there are three general directions which might currently be emphasized in the search for relevant social skills:

further expansion and methodological tightening of the peer validation approach, an emphasis on discovering behavioral changes that might improve the perceived level of social skill among socially inadequate individuals (instead of attempting to relate behaviors in the absolute to the construct of social skill), and an emphasis on idiographic analysis.

The Present Study

In the present study, 42 subjects, both male and female, who reported to be high in social avoidance and distress (Watson & Friend, 1969), each interacted with three different randomly assigned peers of the opposite sex in dyadic social interactions. The 126 peers were selected from the middle range of social avoidance and distress. Peers rated the social competence of subjects and estimated which conversational behaviors should be modified in order for the subjects to appear more socially skilled. Further, several design and measurement practices were based on the assessment literature related to the "halo effect" (Cooper, 1981). These improvements should provide greater confidence in the discriminant validity of the particular ratings which observers made. Generalizability analyses (Cronbach, Gleser, Nanda, & Rajaratnam, 1972) are presented to determine the consistency of peer-nominated suggested changes. Additional analyses of these suggested changes are presented with the intention of clarifying whether nominated changes should be

implemented across subjects or whether individual assessment and treatment seems warranted.

Statement of Specific Hypotheses

1. It was predicted that peers would recommend specific changes in the subject's overt conversational behavior. It was further predicted that independent peers who talked with the same subject would show low, but statistically significant, agreement concerning which changes should be made by the subject.

2. It was predicted that peers, on average, would judge most of the subjects as likely to benefit from a specific set of three or four behavior changes (out of the list of nine behaviors previously discussed). It was further predicted that coherent subgroups of subjects would be discovered who were judged as being likely to benefit from changing different behaviors. Thus, for example, some subjects might be judged as benefiting from no behavior changes, others might benefit from being more positive and showing greater verbal attention, others might benefit from a general increase in talking, and a fourth group might benefit primarily from changes in nonverbal behavior.

CHAPTER II

METHOD

Subject Selection

Subjects were solicited from an introductory psychology course. Students attended a group questionnaire session and completed several personality questionnaires for extra course credit. These questionnaires included the Social Avoidance and Distress scale (Watson & Friend, 1969) (see Appendix A). Students scoring 16 or above on the Social Avoidance and Distress scale were considered to be socially inadequate, while students scoring between 3 and 8, inclusive, were considered to be representative of the normal range. These norms were determined during the first of the two academic terms that this study was conducted and represent approximately the top 10% and middle 40% of this population, respectively. Students scoring in these ranges were phoned by the author and invited to participate in an acquaintanceship study for extra credit. Socially inadequate subjects were told this would involve talking individually with three students for 10 minutes each in a social conversation. The socially adequate peers were told this would involve talking with one student for 10 minutes in a social conversation. Students were randomly assigned to one of the available time blocks, with each socially inadequate subject paired with four peers of the opposite sex.

Alternate times were randomly assigned if the student reported that he/she could not attend the time initially requested. The fourth peer in each block was scheduled to reduce the impact of missed appointments; only three interactions were intended to occur.[1]

Subjects

There were 21 male and 21 female socially inadequate subjects. There were 63 male and 63 female peers. The mean age of each group was as follows: male socially inadequate, M = 19.3, SD = 1.8; female socially inadequate, M = 19.2, SD = 2.0; male peers, M = 19.3, SD = 1.2; female peers, M = 18.6, SD = 0.8. The mean Social Avoidance and Distress score for each group at the time of the initial screening assessment session was: male socially inadequate, M = 19.9, SD = 3.4; female socially inadequate, M = 18.2, SD = 2.0; male peers, M = 5.4, SD = 1.7; female peers, M = 4.6, SD = 1.4. The mean Social Avoidance and Distress score at the time of the social conversation was: male socially inadequate, M = 18.0, SD = 4.6; female socially inadequate, M = 15.3, SD = 3.7; male peers, M = 4.1, SD = 2.8; female peers, M = 3.0, SD = 1.9.

Procedure

Socially inadequate subjects were met by a female research assistant, who obtained their informed consent (see Appendix B) and provided three questionnaires for them to complete: the Social Avoidance and Distress scale, an expectancy rating form designed for

this study (see Appendix C), and the Spielberger State Anxiety Inventory (Spielberger, Gorsuch, & Lushene, 1970) (see Appendix D). The rating form included brief instructions for the social interaction task and four expectancy ratings concerning how socially skilled, socially anxious, likable, and physically attractive the subject expected to seem. These ratings were on 1 to 9 Likert scales. Subjects were asked to interact for 10 minutes with a student of the opposite sex, as if meeting someone for the first time. Subjects were asked not to talk about any previous participation in the study.

While the socially inadequate subject completed these forms, the experimenter left the room and greeted the peer subject, who had been asked to wait in a different area. Peers were given these same forms to complete and were given the same instructions as socially inadequate subjects. After a few minutes, the experimenter collected all forms from both individuals, and then brought the peer subject into the room where the socially inadequate subject had been waiting. The subjects were introduced, and the experimenter left the room. While conversing, the subjects sat across from each other in comfortable chairs, approximately 1.3 meters apart, with a small table in between. The interactions were observed by the research assistant and the author through a one-way window, which presented a 90° side view of both individuals. After 10 minutes, the assistant signaled the end of the conversation and escorted the peer subject back to his/her previous room.

Within their respective rooms, both subjects rated their own performance on the four dimensions previously rated (social skill, anxiety, likability, and physical attractiveness) (see Appendix E), and they completed the Spielberger State Anxiety Inventory. In addition, both subjects rated their partner's behavior on the same four dimensions (see Appendix F).[2] Peers also completed an additional form (see Appendix G) concerning specific suggested changes in their partner's behavior. The two peer rating forms, which concerned the behavior of the socially inadequate partner, constituted the data of interest in this study. The other rating forms and the Spielberger State Anxiety Inventory were included primarily to obscure the purpose of the study and to minimize the possibility that the socially inadequate individuals would be identified as the subjects of this study. The second peer rating form was designed so that peers completed two open-ended questions before they were presented with rating scales concerning specific changes in partner behaviors. The first open-ended question requested the subject to list any suggested changes in his/her partner's conversational style or behavior, while the second question requested any suggested changes that would make his/her partner look more physically attractive. This procedure was intended to minimize demand and halo effects on the specific ratings (cf. Cooper, 1981). Further, the specific ratings included the list of nine behaviors which have each received research support in two or more studies as being relevant to social skill, along with four additional behaviors which have no

research support and were included as halo ratings. The potential social skills were: eye contact; smiles; verbal personal attention; stammers, trembling, or fidgets (hereafter referred to simply as "fidgets"); talking time; hand gestures; pauses; compliments; and positive statements about other people or things (hereafter referred to as "uppers"). The four halo ratings were: criticism, foul language or swearing, self-praise, and extensiveness of vocabulary. Peers rated the degree of change suggested for each behavior, and then they rated the social skill impact of any change they suggested. The degree of change ratings were on 9-point scales where 1 indicated "Should have much less," 5 indicated "Don't change," and 9 indicated "Should have much more." The impact ratings were on 9-point scales where 0 indicated "No change," 4 indicated "Moderate improvement," and 8 indicated "Tremendous improvement." If the halo variables are judged irrelevant, while some of the other behaviors are considered relevant targets, then greater discriminant validity of these behaviors would be established. It was hoped that halo and demand effects might not be problematic in this study because fairly specific behaviors were rated (Cooper, 1981).

While the peer completed the form concerning suggested changes in his/her partner's behavior and appearance, the socially inadequate subject was given a second set of the preconversation instructions, expectancy rating form, and the Spielberger State Anxiety Inventory, in order to prepare him/her for the second conversation. Meanwhile, the second peer subject arrived in a different area and was given the

set of preconversation forms by the assistant and was given the same instructions as the first peer. The second conversation occurred, and postconversation rating forms were completed as discussed before. The third, and final, conversation occurred with a different peer using the same procedures and same pre- and postconversation rating forms as discussed before. Thus, each socially inadequate subject interacted with three different peers of the opposite sex under the same instructional set.

CHAPTER III

RESULTS

Evidence of Random Assignment

Three peers of the opposite sex were randomly assigned each socially inadequate subject. Data available on the peers immediately before they met their conversational partner included: age, Social Avoidance and Distress score, Spielberger State Anxiety score, and ratings of expected social skill, anxiety, likability, and physical attractiveness. Subjects were compared in a one-way MANOVA[3] on their respective peers' responses to these measures, as one indicant of whether random assignment had been successful. Subjects were considered to be a random factor in this analysis. There was no evidence of differential peer characteristics across Subjects, MANOVA $F(287, 551) = 1.02$, $p < .44$.

Suggested Change Rating Scales

Peers completed rating scales indicating the direction and degree of change, if any, that they would recommend in 13 specific categories of their partner's behavior. Nine of the included behaviors had received some support from the research literature as relevant to social skill, while four behaviors had no research support and were

included as comparison variables to investigate the possibility of halo effects.

Discriminant validity. In order to compare whether peers recommended the social skill variables more than the halo variables, an average score across the three peers was computed for each socially inadequate subject on each of the 13 suggested change variables. Each of these scores was then transformed to be the absolute deviation from a score of 5, since 5 represented the "Don't change" anchor on each 9-point scale. Absolute deviation scores were used because it was expected that the direction of suggested change would vary depending on which variable and/or subject was being considered. A dependent t test was conducted comparing the average of the absolute deviation scores for the halo variables with the average for the social skill variables. The social skill variables (mean absolute deviation from 5 = .63) were recommended at a higher level than the halo variables (mean absolute deviation from 5 = .22), $t(41) = 9.04$, $p < .0001$, two-tailed. Moreover, each of nine individual t tests indicated significant differences between the average absolute deviation for the halo variables and the average absolute deviation for each of the nine individual social skill changes. The smallest t value was for fidgets, $t(41) = 3.34$, $p < .0018$, two-tailed. Thus, there is evidence across subjects that each social skill variable was considered to be more relevant for change than the mean of the halo variables.

Generalizability. Analyses were conducted to determine whether the mean rating per subject (across the three peers) on each suggested change variable was reliable. Reliability, in this context, is an index of the consistency of suggested change in a specific behavior, based on three conversations, as perceived by independent peers, each of whom was present for only one conversation with one subject. This design fits readily within an analysis of variance model, where total variability on each scale is partitioned into within-subject variance and between-subject variance. That is, three randomly selected peers were nested within each of 42 randomly selected socially inadequate subjects.[4] The following intraclass correlation is a reliability estimate for the mean rating of three peers: (Mean Square Between Subjects - Mean Square Within Subjects) / Mean Square Between Subjects (Shrout & Fleiss, 1979). Such intraclass correlations have been described by Cronbach et al. (1972) as generalizability coefficients, or G-coefficients. The modernized theory of reliability, named Generalizability Theory, which Cronbach et al. (1972) have presented, has become increasingly popular in the behavioral assessment literature, precisely because it forces one to be specific about the domain of generalization. Generalizability coefficients are presented in Table 2 for each of the suggested change variables and for the mean halo variable. The halo variables were rescaled to be in the same direction before a mean was computed. Means and the standard error of each mean, as computed across the 42 subjects, are also presented in Table 2. The scales for pauses and fidgets were

Table 2

Means and G-coefficients for Suggested Change Ratings

Variable	Mean	STDERR	G-coefficient	Significance Level of G-coefficient[a]
Eye Contact	5.74	.12	.547	$p < .002$
Smiles	5.69	.11	.573	$p < .0005$
Talk Time	5.61	.12	.479	$p < .006$
Hand Gestures	5.56	.11	.459	$p < .009$
Pauses	5.53[b]	.10	.449	$p < .02$
Fidgets	5.51[b]	.10	.514	$p < .003$
Verbal Attention	5.80	.09	.160	$p < .25$
Uppers	5.43	.06	.000[c]	$p < .68$
Compliments	5.39	.05	.000[c]	$p < .66$
Mean Halo	5.21	.02	.057	$p < .41$

[a]As computed by $F(41, 84) = MS$ Subjects/MS Peers(Subjects) (Shrout & Fleiss, 1979). The multivariate test on the set of nine nonhalo variables was: $F(369, 692) = 1.44$, $p < .0001$.

[b]Reverse scaled for comparison purposes. The original mean was an equal amount less than 5.00.

[c]Originally a negative value.

reversed in this presentation to facilitate comparison with the other scales. Six of the nine social skill scales were significantly generalizable, while three were not. The mean halo rating was also nonreliable. The interpretation of these generalizability coefficients is illustrated for the example of the eye contact rating: For one socially inadequate subject selected at random, the mean rating of three peers, also selected randomly, would be expected to share 54.7 percent of the variance with the subject's true eye contact score under similar conditions. The ratings of verbal attention, uppers, and compliments were not reliable, so no further analyses of these variables will be presented. However, the implications of these results will be considered in the Discussion.

Rating specificity. There is no way to insure that raters responded only to the relevant behavioral cues when they made each rating. However, Thorndike (1920) argued that rating intercorrelations that are too high and too even, as compared with real relationships, are evidence that raters have deviated from the intended rating dimensions. Intercorrelations of the peer ratings, averaged across three peers, are shown in Table 3 for the six reliable suggested change scales. These correlations ranged in absolute value from .01 to .65, with a median of .18. They appear to be regularly spaced throughout the range. Only four of the 15 correlations were significantly greater than 0 when using a $p < .05$ criterion. If one adopted a $.05/15 = .0033$ criterion, in order to control for the number of tests conducted, then only the intercorrelation of .65 would

Table 3

Intercorrelations of Reliable Suggested Change Scales

	F	TT	EC	HG	P	S
Fidgets (F)		-.10	-.27	.43*	.16	-.01
Talk Time (TT)			.17	.28	-.26	.32*
Eye Contact (EC)				-.06	-.18	.65**
Hand Gestures (HG)					-.02	-.01
Pauses (P)						-.31*
Smiles (S)						

* $p < .05$, two-tailed

** $p < .0033$, two-tailed

be significantly different from 0. It is expected that there are reliable relationships among some of these behavioral events, since they are all supposed to assess conversational competence. Thus, it does not appear that these rating intercorrelations are particularly high or particularly even. These data are not inconsistent with the possibility that raters attended to the proper cues.

Analysis of hypothesized differential recommended changes. The six reliable suggested change rating scales were analyzed in a two-factor mixed model analysis of variance: 42 levels of the socially inadequate Subject factor x 6 levels of the Measures factor.[5] Subjects was a random factor, while Measures was a fixed repeated factor. The implicit Peer factor was a nested random factor within Subjects, and constituted the experimental units in this design. Thus, a complete specification of effects would be: $Peers_3(Subjects_{42})$ x $Measures_6$. The nesting of the random Peer factor within the random Subject factor allows for tests on the Subject factor and the Subjects x Measures interaction which are not possible in simpler designs. Complete results are shown in Table 4 for these analyses. The scores for fidgets and pauses were reversed prior to this analysis so that all measures were scaled in the same direction. The Greenhouse-Geisser procedure (see Winer, 1971) was used to estimate the degree of departure from the sphericity assumption in the error variance-covariance matrices, and the preferred, more conservative degrees of freedom and probability estimates are provided in Table 4 for those effects which involve the Measures factor. There was significant

Table 4

ANOVA on Suggested Change Ratings

Source	Error Term	MS	df	F	p <
Subjects	P(S)	2.72	41, 84	2.36	.0005
Peers (Subjects) Error		1.16			
Measures	SM	1.05	4.1, 166.1	.83	.52
Subjects x Measures	P(S)M	1.27	188.2, 385.6	1.91	.0001
Peers (Subjects) x Measures Error		.66			

Note: Greenhouse-Geisser Epsilon was .81 when SM was the error term and .92 when P(S)M was the error term.

variation among Subjects, there was no evidence of significant variation for the main effect of Measures, and there was a significant Subjects x Measures interaction.

Variance component estimates were also calculated (viz., Dixon, Brown, Engelman, Frane, Hill, Jennrich, & Toporek, 1981; BMDP8V) as an aid in judging the magnitude of the various effects. Table 5 shows the variance components, the proportions of total variance, and the ratios formed by dividing the variance component for each effect by the sum of itself and all other variance components in its Expected Mean Square. In other words, this last ratio has a more restricted denominator than conventional percent of total variance statistics. This ratio has been advocated by Golding (1975) in favor of the more conventional percent of total variance statistics. By using either set of proportions in Table 5, it appears that the Subjects x Measures interaction is nontrivial. It rivals the strength of the Subject main effect and appears to be of larger magnitude than the non-significant Measures main effect. Together, the Subjects main effect and the Subjects x Measures interaction accounted for 25.3 percent of the total variance, which is an appreciable amount, especially considering the restricted range of this sample.

Having established that the Subjects x Measures interaction is reliable and nontrivial, it is important to better understand the form of this interaction. Obviously, a 42 x 6 interaction poses considerable interpretational difficulty. On a general level, however, the interaction indicates that the relative level of the various

Table 5

Variance Component Analyses on Suggested Change Ratings

Source	Variance Component	Proportion of Total Variance	Variance Component/All Variance Components in EMS
Subjects	.087	.076	.311
Peers (Subjects) Error	.193	.168	
Measures	.000[a]	.000	.000
Subjects x Measures	.202	.177	.234
Peers (Subjects) x Measures Error	.662	.579	

[a]Originally a negative value.

suggested changes differs depending on the subject. In other words, there are some differential suggested changes. As a next step in clarifying the interaction, simple main effect tests on the Measures factor were conducted for each subject. The pooled P(S)M error term and the more conservative Greenhouse-Geisser degrees of freedom were used in these analyses. Results indicated significant variation across measures for 12 of the 42 subjects. The smallest of the significant F's was: $F(4.6, 385.6) = 2.30$, $p < .05$; the largest was: $F(4.6, 385.6) = 10.01$, $p < .0001$. Thus, 12 of the subjects had significantly non-flat profiles. The means for these 12 subjects are shown in Table 6. These means also illustrate that the interaction was clearly disordinal. The rank order of subjects was not the same across measures, nor was the rank order of measures the same across subjects. Post hoc paired mean comparisons could be conducted on these data, but those analyses would not be particularly helpful in determining which behaviors are relevant, and in which direction, for each subject. For example, if a mean of 6.33 on hand gestures was deemed significantly greater than a mean of 4.33 for smiles, then one could infer that hand gestures were judged to need increasing more than smiles were judged to need increasing. However, since 4.33 is below 5, smiles were not judged to need increasing. In this instance, they were judged to need decreasing. The point of this example is that relative comparisons between means are not as

Table 6

Means for 12 Subjects With Differential Suggested Changes[a]

3.67 Hand Gestures	5.00 Talk Time	5.00 Eye Contact	5.00 Smiles	6.00 Pauses	7.00 Fidgets
4.33 Smiles	5.33 Eye Contact	5.33 Pauses	5.67 Talk Time	6.33 Hand Gestures	7.00 Fidgets
4.00 Hand Gestures	4.67 Pauses	5.33 Talk Time	6.00 Fidgets	7.00 Smiles	8.00 Eye Contact
5.00 Talk Time	5.00 Pauses	5.33 Fidgets	6.33 Hand Gestures	6.67 Eye Contact	6.67 Smiles
4.67 Pauses	5.67 Fidgets	5.67 Eye Contact	5.67 Smiles	6.00 Hand Gestures	7.33 Talk Time
5.33 Hand Gestures	5.67 Eye Contact	6.33 Fidgets	6.33 Smiles	6.67 Talk Time	7.33 Pauses
5.33 Fidgets	5.67 Hand Gestures	5.67 Pauses	6.00 Talk Time	6.33 Eye Contact	7.33 Smiles
3.67 Hand Gestures	5.00 Talk Time	5.00 Pauses	5.67 Smiles	6.67 Eye Contact	7.00 Fidgets
5.00 Hand Gestures	5.33 Pauses	6.33 Eye Contact	6.67 Fidgets	7.00 Talk Time	7.33 Smiles
5.00 Fidgets	5.00 Pauses	5.33 Eye Contact	5.33 Smiles	6.67 Hand Gestures	7.00 Talk Time
5.00 Fidgets	6.00 Talk Time	6.00 Hand Gestures	7.00 Pauses	7.33 Smiles	7.67 Eye Contact

Table 6 (Continued)

5.67 Fidgets	5.67 Talk Time	5.67 Pauses	6.33 Smiles	6.67 Hand Gestures	7.67 Eye Contact

[a]Means were underlined if their 99 percent confidence interval does not cross 4.69 or 5.31. The means for fidgets and pauses are reverse-scaled in this presentation.

informative as comparisons of each mean with the "Don't change" midpoint anchor of the scale, or with the absolute mean halo score, which would be somewhat more conservative. The mean absolute deviation from 5 on the halo variables was .31 for these 12 subjects. Thus, 99 percent confidence intervals were computed around the social skill means for each subject.[6] All means that were significantly above 5.31 or significantly below 4.69 were underlined in Table 6. The corresponding behaviors were judged significantly to need changing in these subjects. Two subjects were judged in need of decreasing fidgets, two were judged in need of increasing smiles and eye contact, and the remaining eight subjects each had his/her own idiosyncratic pattern of significant suggested changes.

As previously noted, the other 30 subjects were not shown to have significant differences among their mean scores for the six behaviors. In other words, they had relatively flat profiles. Of these 30 subjects, however, there seems to be an important conceptual distinction between those subjects who had elevated or depressed flat profiles versus those who had flat profiles with a mean near 5. In other words, an elevated flat profile would indicate rather consistent deficits across the six behaviors (chronic deficits as opposed to differential deficits), while a profile average near 5 (or near the mean halo score which was slightly higher) is consistent with the claim that there were no deficits. A depressed profile, if it were to occur, would indicate that on average the subject emits too many of the social skills and/or too few of the behaviors that

were reverse-scaled (fidgets and pauses). Table 7 shows a frequency distribution of the mean suggested change scores for these 30 subjects who had relatively flat profiles. Fidgets and pauses were reverse-scaled before computing these means. There was significant variation on the mean suggested change scores for these 30 subjects, $F(29, 60) = 1.98$, $p < .02$. Ninety-nine percent confidence intervals were computed around each subject's mean score. The P(S) error MS of .161 ($df = 60$) was used in this analysis. Five of the subjects had means significantly greater than the absolute mean halo rating (which was 5.19 for these subjects). The means of these five subjects ranged from 5.83 to 6.44. None of the subjects' means were significantly less than 4.81.

When using the decision rules and variables applied in this study, it appears that out of 42 subjects who reported social avoidance and distress, 12 (28.6%) had differential suggested changes in their overt conversational behavior, 5 (11.9%) had relatively general overt conversational deficits, and 25 (59.5%) were not reliably shown to have any overt conversational deficits. Interestingly, there does not appear to be a difference among these three groups on the mean Social Avoidance and Distress score just prior to the social conversation, $F(2, 39) = 0.07$, $p < .94$. Thus, there seems to be a relevant conceptual difference in the way subject groups were viewed by the peers, but subjects did not view themselves differently as assessed by the Social Avoidance and Distress scale score.

Table 7

Mean Suggested Change Scores for 30 Subjects With Relatively Flat Profiles[a]

Mean	n	Percent	Cumulative Percent
6.44	1	3.3	3.3
6.00	2	6.7	10.0
5.83	2	6.7	16.7
5.72	1	3.3	20.0
5.67	1	3.3	23.3
5.61	2	6.7	30.0
5.50	3	10.0	40.0
5.44	6	20.0	60.0
5.39	3	10.0	70.0
5.33	2	6.7	76.7
5.28	2	6.7	83.3
5.17	1	3.3	86.7
5.11	1	3.3	90.0
5.06	2	6.7	96.7
4.78	1	3.3	100.0

[a]Means were underlined if their confidence interval does not intersect 4.81 or 5.19.

Impact Scales

Peers also rated the degree of improvement, on a 0 to 8 scale, that would result if the subjects were to change their behavior in the way suggested. If no improvements were suggested on a scale, then a score of 0 was automatically assigned. Thus, these scores can be interpreted as an index of how important attention to a particular behavior was judged to be. This scale was included because it seemed possible that some behaviors might change a great deal with little effect, while others might change only a little for the same effect. In a different sense, however, these scales are less informative than the suggested change scales because they ignore the direction of change suggested. Thus, these impact ratings can be considered alternate measures that have some advantages and some disadvantages compared with the degree of suggested change scales.

Discriminant validity. A dependent t test indicated that the average impact rating for the social skill variables (M = 1.87) was greater than for the halo variables (M = 0.66), $t(41) = 9.68$, $p < .0001$, two-tailed. Moreover, individual t tests indicated that each of the nine social skills was judged to have a higher impact than the mean of the halo variables. The smallest of the nine t values was for compliments, $t(41) = 3.75$, $p < .0006$, two-tailed.

Generalizability. Means and G-coefficients are shown in Table 8 for the nine social skill impact ratings and the mean halo impact rating. Five of the social skill ratings were significantly generalizable: smiles, talk time, eye contact, pauses, and fidgets. The mean halo rating was not reliable.

Table 8

Means and G-coefficients for Impact Ratings

Variable	Mean	STDERR	G-coefficient	Significance Level of G-coefficient[a]
Smiles	2.24	.32	.643	$p < .0001$
Talk Time	2.10	.27	.417	$p < .02$
Eye Contact	2.05	.27	.406	$p < .03$
Pauses	1.79	.24	.390	$p < .03$
Fidgets	1.62	.24	.508	$p < .004$
Verbal Attention	2.33	.26	.286	$p < .10$
Hand Gestures	2.17	.22	.272	$p < .12$
Uppers	1.29	.19	.052	$p < .41$
Compliments	1.21	.17	.000[b]	$p < .56$
Mean Halo	0.66	.09	.193	$p < .21$

[a]The multivariate test on the set of nine impact ratings was: $F(369, 692) = 1.38$, $p < .0002$.

[b]Originally a negative value.

Rating specificity. Intercorrelations between each of the five reliable impact ratings are shown in Table 9. The correlations ranged from .05 to .73, with a median of .30. Five of 15 correlations were significant at $p < .05$, while two were significant at the comparison adjusted criterion of $p < .005$. These correlations appear higher and more even than the intercorrelations for the degree of change variables, although only two of these intercorrelations indicate as much as 15% shared variance between impact ratings.

Analysis of possible differential impact. The five reliable impact ratings were analyzed in a two-factor mixed model analysis of variance: 42 levels of the socially inadequate Subject factor x 5 levels of the Measures factor. Subjects was a random factor, while Measures was a fixed repeated factor. The Greenhouse-Geisser estimation procedure was used to calculate more conservative degrees of freedom and probability estimates. As shown in Table 10, there was significant variation among Subjects, no evidence of significant variation among Measures, and a significant Subjects x Measures interaction.

Variance component estimates and relevant ratios were calculated as discussed previously and are shown in Table 11 for the impact ratings. Using either set of proportions, the Subjects x Measures interaction appears nontrivial. In both cases the main effect for Subjects appears stronger than the Subjects x Measures interaction.

The Subjects x Measures interaction was further investigated by conducting simple main effect tests on the Measures factor for each

Table 9

Intercorrelations of Reliable Impact Ratings

	F	TT	EC	P	S
Fidgets (F)		.05	.34*	.34*	.21
Talk Time (TT)			.17	.38*	.46**
Eye Contact (EC)				.18	.73**
Pauses (P)					.26
Smiles (S)					

* $p < .05$, two-tailed

** $p < .005$, two-tailed

Table 10

ANOVA on Impact Ratings

Source	Error Term	MS	df	F	p <
Subjects	P(S)	20.76	41, 84	3.01	.0001
Peers (Subjects) Error		6.90			
Measures	SM	7.87	3.1, 127.3	1.27	.29
Subjects x Measures	P(S)M	6.20	147.5, 302.2	1.51	.0008
Peers (Subjects) x Measures Error		4.10			

Note: Greenhouse-Geisser Epsilon was .78 when SM was the error term and .90 when P(S)M was the error term.

Table 11

Variance Component Analyses on Impact Ratings

Source	Variance Component	Proportion of Total Variance	Variance Component/All Variance Components in EMS
Subjects	.924	.130	.401
Peers (Subjects) Error	1.380	.194	
Measures	.013	.002	.003
Subjects x Measures	.700	.098	.146
Peers (Subjects) x Measures Error	4.099	.576	

subject. Nine subjects had significant differences ($p < .05$) among the impact ratings. Means for these nine subjects are shown in Table 12. The mean halo impact rating was 0.92 for these subjects. Ninety-nine percent confidence intervals were computed around each subject's mean impact scores. Means that were reliably above 0.92 were underlined in Table 12. For two subjects it was judged that modifying eye contact and smiles would result in significant improvement. Changes in fidgets were rated as likely to have significant impact in three subjects, changes in smiles were considered to have significant impact for two subjects, while the remaining subjects each had his/her own idiosyncratic pattern of significant impact ratings.

Mean impact ratings for the 33 subjects with relatively flat profiles are shown in Table 13. There was significant variation on these means across subjects, $F(32, 66) = 3.03$, $p < .0001$. Ninety-nine percent confidence intervals were computed around each mean, and those means that were significantly above the mean halo impact rating of 0.60 (computed for these subjects) were underlined in Table 13. Ten subjects had means which were significantly above the mean halo impact level.

Thus, when using the decision rules and variables applied in this study, the suggested changes were viewed as having differential impact for nine (21.4%) subjects, a general positive impact for 10 (23.8%) subjects, and 23 (54.8%) subjects were not judged to improve from altering their conversational behavior. In contrast to the

Table 12

Means for Nine Subjects With Differential Impact Ratings[a]

0.00 Fidgets	0.00 Pauses	1.00 Smiles	2.00 Eye Contact	5.67 Talk Time
0.00 Talk Time	0.00 Eye Contact	0.00 Smiles	2.00 Pauses	4.67 Fidgets
1.33 Fidgets	1.33 Pauses	3.67 Talk Time	4.00 Eye Contact	6.33 Smiles
1.67 Pauses	4.00 Fidgets	4.00 Talk Time	4.00 Eye Contact	7.00 Smiles
0.00 Talk Time	0.00 Pauses	1.00 Fidgets	4.67 Smiles	5.67 Eye Contact
1.67 Pauses	2.33 Fidgets	2.33 Talk Time	6.33 Smiles	6.67 Eye Contact
0.00 Talk Time	0.00 Smiles	2.33 Eye Contact	2.33 Pauses	4.67 Fidgets
0.67 Pauses	1.33 Smiles	1.67 Talk Time	2.67 Eye Contact	5.67 Fidgets
0.33 Fidgets	1.33 Eye Contact	2.33 Smiles	3.00 Talk Time	5.67 Pauses

[a]Means were underlined if their 99 percent confidence interval does not cross 0.92.

Table 13

Mean Impact Ratings for 33 Subjects With Relatively Flat Profiles[a]

Mean	n	Percent	Cumulative Percent
4.47	1	3.0	3.0
4.27	1	3.0	6.1
3.87	1	3.0	9.1
3.73	1	3.0	12.1
3.13	2	6.1	18.2
2.87	1	3.0	21.2
2.60	1	3.0	24.2
2.47	2	6.1	30.3
2.27	1	3.0	33.3
2.07	1	3.0	36.4
1.87	1	3.0	39.4
1.80	1	3.0	42.4
1.33	2	6.1	48.5
1.27	2	6.1	54.5
1.20	2	6.1	60.6
1.13	2	6.1	66.7
1.07	2	6.1	72.7
1.00	1	3.0	75.8
0.93	1	3.0	78.8
0.87	1	3.0	81.8
0.80	1	3.0	84.8
0.67	1	3.0	87.9
0.60	1	3.0	90.9
0.47	1	3.0	93.9
0.40	1	3.0	97.0
0.00	1	3.0	100.0

[a]Means were underlined if their confidence interval does not intersect 0.60.

groups which were formed based on the degree of change ratings, there does appear to be a significant difference among these groups on the mean Social Avoidance and Distress scores, $F(2, 39) = 4.05$, $p < .03$. Omega squared indicated that these group differences accounted for an estimated 12.7% of the variance in SAD scores. Follow-up t tests indicated that the differential impact subjects viewed themselves as less socially avoidant and distressed ($M = 14.8$) than the general positive impact subjects ($M = 19.7$), $t(39) = -2.63$, $p < .013$. The subjects with no established social skill impact from suggested changes ($M = 16.0$) also reported less social avoidance and distress than the general positive impact subjects, $t(39) = -2.40$, $p < .022$. The differential impact and no established impact subjects did not differ significantly from each other ($p > .10$).

As can be seen from Table 14, there was considerable overlap between the subject classifications formed by the degree of change scales, as compared with the classifications from the impact ratings. Thirty-one of 42 subjects (73.8%) were classified the same using either the degree of change scales or the impact ratings. Relatively more subjects were classified as general problem subjects under the impact ratings. This presumably reflects the fact that the impact ratings do not reflect directional information. There were also a few subjects with no established changes based on the degree of change scales, who were considered to be in either of the other categories based on the impact ratings.

Table 14

Number of Subjects in Each Classification

	Differential Change	General Change	No Change
Differential Impact	7 (16.7%)	1 (2.4%)	1 (2.4%)
General Impact	4 (9.5%)	3 (7.1%)	3 (7.1%)
No Impact	1 (2.4%)	1 (2.4%)	21 (50.0%)

Global Peer Ratings

Generalizability. After the social conversation, peers rated the subjects on four global 9-point scales: social skill, anxiety, likability, and physical attractiveness. Generalizability coefficients were computed for each rating in order to determine whether peers could reliably differentiate subjects, even though peer judgments were based on different conversations. The G-coefficients calculated for the mean rating of three peers were: social skill, .377, $p < .04$; anxiety, .253, $p < .13$; likability, .118, $p < .31$; and physical attractiveness, .713, $p < .0001$. The multivariate test on the set of ratings was also significant, $F(164, 325) = 1.53$, $p < .0006$. No further analyses of the anxiety and likability ratings will be provided due to their apparent lack of generalizability. The correlation between the mean social skill and mean physical attractiveness rating was .493, $p < .0009$, two-tailed, $n = 42$.

Corroboration of suggested change group classifications. If the subject grouping procedures based on the degree of suggested change ratings were meaningful, then one would expect significant variation among the three groups on the peers' global social skill ratings. It is acknowledged that the peer global ratings were not independent of the suggested change ratings used to classify subjects into groups, since they were derived from the same peers. However, such a comparison does seem to provide some corroborative evidence of group differences, and perhaps more importantly, it provides an index of percent of variance attributable to group classification. A one-way

ANOVA across groups on the mean peer social skill rating indicated significant differences among the three groups, $F(2, 39) = 8.45$, $p < .0009$, where 26.2% of the variance can be accounted for by group classification (omega squared, Hays, 1981, p. 349). The mean social skill rating for the general change subjects ($M = 5.07$) was significantly lower than the mean for subjects with no established suggested changes ($M = 6.27$), $t(39) = -3.30$, $p < .003$; the differential change subjects ($M = 5.44$) also received lower ratings of social skill than the no-change subjects, $t(39) = -3.15$, $p < .004$; while there was no reliable difference between the perceived social skill level of the differential change versus general change subjects ($p > .10$).

Analyses were also conducted on the physical attractiveness rating. The ANOVA indicated significant differences among groups, $F(2, 39) = 8.56$, $p < .0008$. Group classification could account for 26.5 percent of the variance in peers' ratings of attractiveness. The differential change subjects ($M = 4.5$) were seen as less attractive than the no-change subjects ($M = 5.85$), $t(39) = -4.13$, $p < .0002$, while the general change subjects ($M = 5.33$) were not significantly different from either group ($p > .10$ for each comparison).

Corroboration of group classifications from impact ratings. The subject groupings derived from the impact ratings were also compared on the global peer ratings of social skill and physical attractiveness. There was a significant difference among groups on the social skill rating, $F(2, 39) = 4.48$, $p < .02$. Group classification accounted for an estimated 14.2% of the variance. Follow-up tests indicated that the general positive impact subjects ($M = 5.27$) were rated lower in social

skill than the no established social skill impact subjects (M = 6.17), $t(39) = -2.99$, $p < .005$. Neither of the two other comparisons was statistically significant ($p > .10$). The mean peer social skill rating for the differential impact subjects was intermediate between the two other groups (M = 5.85).

There was no reliable difference between groups on the physical attractiveness rating, $F(2, 39) = 1.47$, $p < .25$.

Open-Ended Questions

Social skill suggested changes. Peers were asked to list any specific changes in their partner's conversational style or behavior that would make them seem more socially skilled. It was hoped that requiring peers to complete this open-ended question, and a similar question concerning physical attractiveness, would minimize halo effects on the rating scales which peers subsequently completed (Cooper, 1981). Data from these open-ended questions were also of interest in their own right, but it was expected that it would be difficult to interpret meaningfully these data, given the psychometric limitations of open-ended questions. For four subjects (9.5%), all female, none of the three peers suggested any changes. For 22 of the subjects (52.4%), 11 male and 11 female, there was at least one suggestion by one peer, although no instance in which two or more peers suggested the same change. These data are not presented because of the dubious value of these suggestions considering that there was no agreement. For the remaining 16 subjects (38.1%), 10 male and 6 female, the instances of agreement were as follows:

three male subjects were judged by two peers as likely to seem more socially skilled if they asked more questions. Two male and one female subjects were judged by two peers as benefiting from more eye contact, and an additional male subject was recommended as needing more eye contact by all three peers. Two female and one male subject were judged by two peers as benefiting from talking more, and one additional male subject was so judged by three peers. For one female subject, it was recommended by all three peers that she talk less. It was recommended by two peers that one female subject talk about more personal matters. One male subject was judged by two peers as likely to seem more socially skilled if he had more eye contact and talked more. It was suggested by two peers that one female subject act more enthusiastic. An additional male subject was judged by all three peers as likely to benefit from more eye contact and was also judged by two of the peers as likely to benefit from talking more.

Physical attractiveness suggested changes. Peers were asked to list any specific recommendations that would make their partner look more physically attractive. For nine of the subjects (21.4%), two male and seven female, none of the three peers gave any specific suggestions. For 20 of the subjects (47.6%), 10 male and 10 female, there was at least one suggestion made by one peer, although there were no instances where two or more of the peers listed the same suggestion. Data are not presented for these subjects because of the lack of agreement. For the remaining 13 subjects (31.0%), nine male and four female, the instances of agreement were as follows: three

male and two female subjects were judged by two of three peers as likely to benefit from a change in hairstyle. Two additional male subjects were judged by all three peers as likely to look more attractive if they changed their hairstyle. One male subject was judged by two peers as likely to look more attractive if he changed his hairstyle, shaved off his facial hair, and dressed differently. One male subject was judged to need losing weight by two peers. An additional female subject was judged as needing to lose weight by all three peers and was judged as likely to benefit from a change in hairstyle by two of the peers. One male subject was judged by two peers as likely to benefit from shaving off facial hair, dressing differently, and improving posture. One male subject was judged by two peers as likely to benefit from a change in hairstyle and from getting contact lenses or different glasses. An additional female subject was judged by two peers as likely to look more attractive if she improved her posture.

CHAPTER IV

DISCUSSION

Findings and Implications

The results of this study strongly suggest the importance of individualized assessment and treatment of social inadequacy. Although all subjects reported a significant degree of social avoidance and distress, problematic overt conversational behavior was reliably indicated by the peers in only 40-45% of the subjects. Moreover, for those subjects who were judged to have problematic conversational behavior, differential changes seemed warranted somewhere between 50-65% of the time.

One encouraging aspect of this research is that peers generally validated the relevance of many of the behaviors that have somewhat successfully been investigated in prior research (cf. Conger et al., 1980). Most of the instances of agreement in the open-ended social skill question concerned variables with some prior research support. Moreover, on the specific ratings, each of the nine behaviors was rated significantly higher than the mean halo variable for the degree of change scales and for the impact ratings.

Six of nine degree of change scales (eye contact, smiles, talk time, hand gestures, pauses, and fidgets) and five of nine impact scales (eye contact, smiles, talk time, pauses, and fidgets) were

significantly reliable. Moreover, the multivariate test indicated significant reliability for each set of nine ratings. Most of the G-coefficients for the behavioral ratings were in the .5 range. This is a reasonably high value, considering that each peer based his/her judgments on a different conversation. In other words, it is not reasonable to expect these values to be as high as an observer agreement correlation where observers watched the same conversation. In this regard, the relative standardness of physical appearance is the obvious explanation why the G-coefficient for the global attractiveness rating was the highest of any rating in this study. All G-coefficients were computed for the mean across three peers. The G-coefficients would almost surely have been larger if more peers interacted with each subject, although subject fatigue could readily become a problem. The use of four or more peers would also improve the power of the follow-up comparisons, since the n used in this study to compute the standard error of each mean was 3.

The implications of this research for the three nonsignificantly generalizable ratings remains largely unclear: verbal personal attention, compliments, and uppers. These three behaviors were nonreliable for both the degree of change and impact ratings, although in both cases all three behaviors were rated significantly higher than the mean halo rating. The mean for verbal personal attention was actually the highest of any behavior on both the degree of change scale and the impact rating. Thus, it might be that verbal personal attention behaviors are often relevant treatment targets, but the

category was simply too broad to permit reliable peer judgments of individual subjects. The category was based on the work of Kupke and colleagues (Kupke, Calhoun, & Hobbs, 1979; Kupke, Hobbs, & Cheney, 1979), but it might be desirable to break this category into finer gradations in order to facilitate peer agreement. Questions, for example, were nominated several times in the open-ended section and were one of the major examples of the Kupke code. The means for compliments and uppers were the lowest of the social skill codes, so it appears that these codes have less peer validational support, based on this one study. As suggested previously, compliments may simply be too difficult a skill for these subjects to implement without prior attention to other behaviors. The relatively low ratings and the lack of reliability for uppers were both surprising.

Both degree of change scales and impact scales were completed for each behavior since it seemed possible that there might be differential impact for a given amount of change, depending on the behavior. As it turned out, the results were fairly similar for both sets of ratings. However, there appeared to be somewhat more evidence for the importance of differential attention on the degree of change scales than on the impact scales. This probably reflects the fact that the degree of change scales included directional information (more or less of a behavior) while the impact scales did not.

It appears that the concern which prompted the inclusion of impact ratings was unfounded. This need not mean that the relationship

between frequency and impact is the same for all of these behaviors, however. The preferred explanation may well be that the terms "more" and "less" were abstract enough that peers made adjustments for the base rates of these behaviors. Thus, "much less" pauses may refer to two or three pauses lasting 5-10 seconds each, while "much less" fidgets may refer to a large number of small events that occurred over several minutes of the conversation. In other words, the peers seem to have abstracted the intended meaning of the degree of change scales, instead of treating the scales legalistically. If true, then the impact scales would be unnecessary and would, in fact, be less informative since the scales do not include directional information. The concern, put bluntly, that prompted the inclusion of the impact scales was that there was no way to tell a priori whether peers would rate each behavioral dimension on the degree of change scales as if they were all on the same conceptual scale, despite the fact that the ratings all had the same anchors. The impact scale was sufficiently concrete that there was no psychometric concern about comparing ratings for different behaviors, although, unfortunately, impact ratings do not provide information about the direction of suggested change. It would be informative to study the relationships among impact ratings, degree of change ratings, and the frequency or duration of behavioral events. Without the availability of these data, it seems that the degree of change ratings provide more important information since they include the directional component. The degree of change ratings also appeared to be somewhat more generalizable and had lower

intercorrelations than the impact ratings. Moreover, the group classifications determined by the degree of change scales accounted for almost twice as much variance in global peer ratings of social skill, as compared with the classifications that were derived from the impact ratings.

The results of this study may also clarify the somewhat ambiguous information provided by the 18 or more studies which attempted to identify behavioral components of social skill. Mean group comparisons obscure individual differences, and it appears that individual variation is quite large within this population. There were even some subjects where the average peer ratings on the degree of change scales indicated that the subject should have less talk time, less eye contact, fewer hand gestures, more pauses, less smiling, or more fidgets. None of these instances was demonstrated to be a reliable effect based on the confidence interval analyses, although as previously noted, the confidence intervals were particularly broad given the n of 3 used to compute the standard error of each mean. On the open-ended social skill question, one subject was evaluated by all three peers as talking too much. Large individual variation may have contributed to the somewhat inconclusive results of prior group assessment studies.

Most of the prior assessment studies have used analogue populations similar to the present study, although the selection criterion of 16 on the SAD which was used in this study seems more rigorous than most of the various selection criteria which have previously

been employed. Research has not clarified whether specific conversational problems are different for clinical versus analogue groups, although the usefulness of analogue research for the target problem of social inadequacy has generally been acknowledged (Borkovec et al., 1974).

This study has also provided what appear to be the first assessment data on specific problematic physical characteristics of some socially inadequate individuals. While it has long been recognized that physical characteristics profoundly affect interpersonal attraction (Dion, Berscheid, & Walster, 1972), clinical assessment and treatment research has not adequately emphasized the role of physical attractiveness for the problem construct of social inadequacy (Arkowitz, 1981; Greenwald, 1977). Many aspects of one's physical appearance are quite modifiable, and attention to some of these characteristics in the context of a comprehensive treatment program seems warranted for some individuals. In this study, at least two peers agreed on one or more changes in physical appearance for 13 of the 42 subjects (31.0%). Hairstyle change was recommended for 10 subjects, shaving off facial hair was recommended for two male subjects, change of dress was recommended for two subjects, losing weight was suggested twice, improving posture was suggested twice, and one subject was judged as likely to benefit from getting contacts or different glasses.

This study has introduced an assessment methodology for identifying problematic conversational behaviors of individual subjects

which meets at least minimum standards of internal validity and may offer substantial improvement in terms of external validity. As noted previously, there is no proof that peers were accurate in their judgments. Instead, a series of convergent analyses and logical arguments were offered which suggest that there is no obvious reason to doubt these judgments. There are limitations to any methodology, and this method was offered as convergent evidence for the relevance of behaviors which already had moderate research support from more conventional methodologies.

It also should be emphasized that this study focused on the behavior and physical appearance of individuals who reported high social avoidance and distress. Other individuals might have different nominated changes, in terms of topography and/or degree of change. No high socially competent group was included in this design because the emphasis was on using peers to validate the clinical relevance of previously nominated conversational behaviors among subjects who approximated likely candidates for therapeutic attention.

This study examined social skills in the context of introductory conversations. More intimate relationships may require different skills. In addition, knowledge of one's conversational partner may idiosyncratically influence the way specific behaviors are interpreted. It is unclear whether the results of this study are relevant for other kinds of conversations besides introductory meetings. In this regard, some of the subjects in this study who had no demonstrated overt behavioral problems in introductory conversations may have overt problems in other settings.

Suggestions for Future Research

Future assessment research employing this method might benefit from attempting to clarify the possibility of gender effects. All behaviors that received peer validational support in this study were examined in the context of cross-sex conversations. It is unknown whether same-sex conversations would have provided similar results. It was also noted parenthetically in footnote #5 that there was some indication of gender effects in this study, although the research design did not permit the unambiguous examination of gender effects.

An additional avenue of future research would be to code independently the frequencies of the investigated subject behaviors and to then study the relationships between the peers' impressions of needed changes and the independently coded behavioral events. Peers may derive their impressions from observed behavior in complex ways. For example, the first minute of the conversation may be weighted more heavily than the remaining minutes. Some attention should also be focused on the quality and timing of conversation skills, and not just their frequency. A peer recommendation that there should be more or less of a certain behavior may be influenced by these other factors and may not just imply that the frequency of the behavioral event should be modified.

A third major expansion on this assessment methodology would involve determining the extent to which clinicians could agree about the relevance of targeting certain social behaviors in individual clients. Moreover, it would be interesting to study the relationships

between peer nominations and therapist nominations. It is unclear whether a therapist should overrule peer nominations if there is disagreement, since one of the primary goals of social skills training is to increase peer acceptance of the client. One of the long-term goals of this assessment methodology would be to determine how clinicians might best use the peer nominations in the context of other assessment information.

After sufficient efforts to improve this assessment method, a strong test of this methodology could be conducted by crossing subjects in each of the assessment groups (differential suggested changes, general changes, and no established changes) with three kinds of therapeutic procedures: individually tailored skills training, packaged skills training, and a third treatment such as cognitive restructuring. Presumably, subjects who report social inadequacy but have no demonstrated conversational problems will improve more, on average, from the nonskills treatment. Moreover, subjects with differential skills problems should benefit more from a differential skills program than from a packaged program. The other mean comparisons would be difficult to predict.

The uniform application of a standard social skills program does not seem warranted for all clients who report problems of social avoidance and distress. While it appears that few authors would recommend such standardization in the day-to-day treatment of real clients in applied settings, it seems inconsistent that the research used to evaluate clinical procedures almost always assumes such standardization.

In this author's opinion, there are two primary reasons why individual differences have been ignored in treatment research. First, analysis of variance has received almost exclusive emphasis as the statistical procedure of choice in academic psychology. This technique was developed by Fisher in the context of agricultural research, where individual variation was appropriately considered to be error variance. Human beings, on the other hand, are much more complex than turnips. Thus, the more interesting analyses of variance for academic psychology include more complex models, instead of the relatively simple models that have been employed. In this study, the random Peer factor was nested within the random Subject factor, which permitted tests on the Subjects main effect and the Subjects x Measures interaction. There is nothing novel or original about this design or analysis. It just represents a somewhat more complex model than has traditionally been employed. It is conceptually analogous to including a Confederate factor in a design which permits tests for Confederate main effects and interactions.

A second explanation for the hesitance of researchers to account for individual differences is that subject classification procedures necessarily involve subjective decisions. For example, the subject with the flattest profile in the differential change group may be only trivially more suited (or, in fact, not reliably more suited) for differential attention than the subject with the bumpiest profile in the general change group. These subject groupings were intended primarily to illustrate the general meaning and thrust of the Subjects x

Measures interaction, rather than to reify any one subject's apparent problems. The point remains, however, that the clinician must treat individual clients who report social avoidance and distress. The kinds of subjective decisions which the applied researcher would prefer to avoid are precisely the same classification issues which must be managed by the clinician.

Undoubtedly, the clinician should be more concerned about Type II errors than the basic researcher. The behavioral model of clinical treatment favors the inclusion of almost any independent variable which is likely to remediate any relevant problem. Thus, the clinician might include some form of social skills training within an individualized treatment package if there was only moderate evidence for problematic overt conversational behavior. There also may be non-specific treatment effects from a social skills program even if there were no apparent overt conversational problems. Thus, if an applied clinician or a treatment study researcher were to adopt some form of this methodology to identify treatment targets, then it would be recommended that he/she be somewhat more liberal in identifying targets, as compared with the interpretation of the data presented in this study. For example, confidence intervals might be computed using a $p < .05$ or $p < .10$ criterion. Alternatively, if future research corroborates the relatively weak impact of halo and demand effects on this type of rating, then means might be interpreted as relevant if the confidence interval does not intersect the "Don't change" midpoint anchor, instead of the absolute mean halo level. In this regard, it

should be pointed out that some or all of the behaviors considered to be halo variables for the purpose of this study may be relevant treatment targets for a small minority of subjects. The mean of these variables was used as the relevant comparison standard instead of treating the 5.00 index literally, simply because there was no prior experience in the use of this type of scale. Replication is needed to corroborate these findings, but it does not appear that halo or demand effects are particularly problematic in this context.

Summary

The major hypotheses of this research were generally supported, although consistent groupings of the differential change subjects were not found. At this point in time, it appears that the targeted behaviors of the differential change subjects are fairly idiosyncratic within the domain of the rating scales and do not fall into consistent categories such as: nonverbal behavior, verbal behavior, and reinforcement processes. Nor can the data yet be interpreted in terms of an integrative model which emphasizes the relationships among physical appearance, verbal behavior, and nonverbal behavior. It is unclear why certain behaviors were recommended for specific subjects. The more general hypothesis of this study received clear support: individual assessment and prescriptive treatment of social inadequacy seems warranted.

REFERENCE NOTE

1. Dow, M. G., Biglan, A., & Glaser, S. R. Multimethod assessment of socially anxious and socially nonanxious women. Manuscript submitted for publication, 1983.

REFERENCES

Argyle, M., & Kendon, A. The experimental analysis of social performance. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 3). New York: Academic Press, 1967.

Arkowitz, H. The assessment of social skills. In M. Hersen & A. Bellack (Eds.), Behavioral assessment: A practical handbook. New York: Pergamon, 1981.

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

Barlow, D. H., Abel, G. G., Blanchard, E. B., Bristow, A. R., & Young, L. D. A heterosocial skills behavior checklist for males. Behavior Therapy, 1977, 8, 229-239.

Bayes, M. A. Behavioral cues of interpersonal warmth. Journal of Consulting and Clinical Psychology, 1972, 39, 333-339.

Bellack, A. S. Behavioral assessment of social skills. In A. S. Bellack & M. Hersen (Eds.), Research and practice in social skills training. New York: Plenum, 1979.

Borkovec, T. D., Fleischmann, D. J., & Caputo, J. A. The measurement of anxiety in an analogue social situation. Journal of Consulting and Clinical Psychology, 1973, 41, 157-161.

Borkovec, T. D., Stone, N. M., O'Brien, G. T., & Kaloupek, D. G. Evaluation of a clinically relevant target behavior for analog outcome research. Behavior Therapy, 1974, 5, 503-513.

Bryant, B., & Trower, P. E. Social difficulty in a student sample. British Journal of Educational Psychology, 1974, 44, 13-21.

Bryant, B., Trower, P., Yardley, K., Urbieta, H., & Letemendia, F. J. J. A survey of social inadequacy among psychiatric outpatients. Psychological Medicine, 1976, 6, 101-112.

Cantor, N., & Mischel, W. Prototypicality and personality: Effects on free recall and personality impressions. Journal of Research in Personality, 1979, 13, 187-205. (a)

Cantor, N., & Mischel, W. Prototypes in person perception. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 12). New York: Academic Press, 1979. (b)

Cherulnik, P. D., Neely, W. T., Flanagan, M., & Zachau, M. Social skill and visual interaction. The Journal of Social Psychology, 1978, 104, 263-270.

Conger, A. J., Wallander, J. L., Mariotto, M. J., & Ward, D. Peer judgments of heterosexual-social anxiety and skill: What do they pay attention to anyhow? Behavioral Assessment, 1980, 2, 243-259.

Conger, J. C., & Farrell, A. D. Behavioral components of heterosocial skills. Behavior Therapy, 1981, 12, 41-55.

Conger, J. C., & Keane, S. P. Social skills intervention in the treatment of isolated or withdrawn children. Psychological Bulletin, 1981, 90, 478-495.

Cooper, A., Furst, J. B., & Bridger, W. H. A brief commentary on the usefulness of studying fear of snakes. Journal of Abnormal Psychology, 1969, 74, 413-414.

Cooper, W. H. Ubiquitous halo. Psychological Bulletin, 1981, 90, 218-244.

Craighead, W. E., Kazdin, A. E., & Mahoney, M. J. (Eds.). Behavior modification: Principles, issues, and applications. Boston: Houghton Mifflin, 1981.

Cronbach, L. J. The two disciplines of scientific psychology. American Psychologist, 1957, 12, 671-684.

Cronbach, L. J. Beyond the two disciplines of scientific psychology. American Psychologist, 1975, 30, 116-127.

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

Curran, J. P. Skills training as an approach to the treatment of heterosexual-social anxiety: A review. Psychological Bulletin, 1977, 84, 149-157.

Curran, J. P. Social skills: Methodological issues and future directions. In A. S. Bellack & M. Hersen (Eds.), Research and practice in social skills training. New York: Plenum, 1979.

Curran, J. P., Miller, I. W., III, Zwick, W. R., Monti, P. M., & Stout, R. L. The socially inadequate patient: Incidence rate, demographic and clinical features, and hospital and posthospital functioning. Journal of Consulting and Clinical Psychology, 1980, 48, 375-382.

Curran, J. P., & Wessberg, H. W. Assessment of social inadequacy. In D. H. Barlow (Ed.), Behavioral assessment of adult disorders. New York: Guilford, 1981.

Dion, K., Berscheid, E., & Walster, E. What is beautiful is good. Journal of Personality and Social Psychology, 1972, 24, 285-290.

Dixon, W. J., Brown, M. B., Engelman, L., Frane, J. W., Hill, M. A., Jennrich, R. I., & Toporek, J. D. (Eds.). BMDP statistical software 1981. Berkeley: University of California Press, 1981.

Dow, M. G., & Craighead, W. E. Cognition and social inadequacy: Relevance in clinical populations. In P. Trower (Ed.), Cognitive approaches to social skills training. Oxford: Pergamon, in press.

Dow, M. G., Glaser, S. R., & Biglan, A. The relevance of specific conversational behaviors to ratings of social skill: An experimental analysis. Journal of Behavioral Assessment, 1981, 3, 233-242.

Ellsworth, P. C., & Carlsmith, J. M. Effects of eye contact and verbal content on affective response to a dyadic interaction. Journal of Personality and Social Psychology, 1968, 10, 15-20.

Galassi, J. P., & Galassi, M. D. Modification of heterosocial skills deficits. In A. S. Bellack & M. Hersen (Ed.), Research and practice in social skills training. New York: Plenum, 1979.

Glasgow, R. E., & Arkowitz, H. The behavioral assessment of male and female social competence in dyadic heterosexual interactions. Behavior Therapy, 1975, 6, 488-498.

Goldfried, M. R., & D'Zurilla, T. J. A behavioral-analytic model for assessing competence. In C. D. Spielberger (Ed.), Current topics in clinical and community psychology. New York: Academic Press, 1969.

Golding, S. L. Flies in the ointment: Methodological problems in the analysis of the percentage of variance due to persons and situations. Psychological Bulletin, 1975, 82, 278-288.

Gottman, J. M. Marital interaction: Experimental investigations. New York: Academic Press, 1980.

Greenwald, D. P. The behavioral assessment of differences in social skill and social anxiety in female college students. Behavior Therapy, 1977, 8, 925-937.

Hays, W. L. Statistics. New York: Holt, Rinehart and Winston, 1981.

Hersen, M. Modification of skill deficits in psychiatric patients. In A. S. Bellack & M. Hersen (Eds.), Research and practice in social skills training. New York: Plenum, 1979.

Jaremko, M. E., Myers, E. J., Daner, S., Moore, S., & Allin, J. Differences in daters: Effects of sex, dating frequency, and dating frequency of partner. Behavioral Assessment, 1982, 4, 307-316.

Kanfer, F. H., & Saslow, G. Behavioral analysis: An alternative to diagnostic classification. Archives of General Psychiatry, 1965, 12, 529-538.

Kazdin, A. Situational specificity: The two-edged sword of behavioral assessment. Behavioral Assessment, 1979, 1, 57-75.

Kiesler, D. J. Some myths of psychotherapy research and the search for a paradigm. Psychological Bulletin, 1966, 65, 110-136.

Kupke, T. E., Calhoun, K. S., & Hobbs, S. A. Selection of heterosocial skills. II. Experimental validity. Behavior Therapy, 1979, 10, 336-346.

Kupke, T. E., Hobbs, S. A., & Cheney, T. H. Selection of heterosocial skills. I. Criterion-related validity. Behavior Therapy, 1979, 10, 327-335.

Lewinsohn, P. M. A behavioral approach to depression. In R. J. Friedman & M. M. Katz (Eds.), The psychology of depression: Contemporary theory and research. New York: Wiley, 1974.

Lewinsohn, P. M., Weinstein, M. S., & Shaw, D. A. Depression: A clinical research approach. In R. D. Rubin & C. M. Frank (Eds.), Advances in behavior therapy--1968. New York: Academic Press, 1969.

Libet, J. M., & Lewinsohn, P. M. Concept of social skill with special reference to the behavior of depressed persons. Journal of Consulting and Clinical Psychology, 1973, 40, 304-312.

Marlatt, G. A., & Donovan, D. M. Alcoholism and drug dependence: Cognitive social learning factors in addictive behaviors. In W. E. Craighead, A. E. Kazdin, & M. J. Mahoney (Eds.), Behavior modification: Principles, issues, and applications. Boston: Houghton Mifflin, 1981.

Martinez-Diaz, J. A., & Edelstein, B. A. Heterosocial competence: Predictive and construct validity. Behavior Modification, 1980, 4, 115-129.

Marzillier, J. S., & Winter, K. Success and failure in social skills training: Individual differences. Behaviour Research and Therapy, 1978, 16, 67-84.

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.

Shrout, P. E., & Fleiss, J. L. Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 1979, 86, 420-428.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. Manual for the strait-trait anxiety inventory. Palo Alto, California: Consulting Psychologists' Press, 1970.

Sullivan, H. S. The interpersonal theory of psychiatry. New York: Norton, 1953.

Thorndike, E. L. A constant error in psychological ratings. Journal of Applied Psychology, 1920, 4, 25-29.

Trower, P. Situational analysis of the components and processes of behavior of socially skilled and unskilled patients. Journal of Consulting and Clinical Psychology, 1980, 48, 327-339.

Twentyman, C. T., & McFall, R. M. Behavioral training of social skills in shy males. Journal of Consulting and Clinical Psychology, 1975, 43, 384-395.

Twentyman, C. T., & Zimering, R. T. Behavioral training of social skills: A critical review. In M. Hersen, R. M. Eisler, & P. M. Miller (Eds.), Progress in behavior modification (Vol. 7). New York: Academic Press, 1979.

Watson, D., & Friend, R. Measurement of social-evaluative anxiety. Journal of Consulting and Clinical Psychology, 1969, 33, 448-457.

Winer, B. J. Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill, 1971.

Youngren, M. A., & Lewinsohn, P. M. The functional relation between depression and problematic interpersonal behavior. Journal of Abnormal Psychology, 1980, 89, 333-341.

Zigler, E., & Phillips, L. Social competence and outcome in psychiatric disorder. Journal of Abnormal and Social Psychology, 1961, 63, 264-271.

FOOTNOTES

[1]Excluding seven students (out of 291 contacted) who stated that they had dropped the course or had already earned the maximum amount of extra credit, and excluding 24 students who agreed to participate but could not be scheduled in the remaining times available, the following percentages indicate the proportion of students contacted in each group who both agreed to participate and showed up for the experiment: male SI, 89%; female SI, 92%; male peer, 91%; female peer, 83%. Subjects were run between 3 and 20 days after attending the group questionnaire session.

[2]Peers and socially inadequate subjects also indicated on this form whether they had previously known the person they spoke with. Out of 126 conversations, in only one instance had the interactants met before, and both indicated that they did not know each other very well. In one additional instance, the interactants said they had seen each other on campus before, but had never met.

[3]The F test derived from Wilk's lambda was used as the test statistic in all MANOVA analyses.

[4]A more complex model could have been used, such as: $Peer_3(Subject_{21}(Gender_2))$. In this case, the Gender factor would be a dummy variable representing male subjects talking to females, versus female subjects talking to males. The MANOVA on the nine social skill variables using this model indicated no Gender effects, $F(9, 32) = 1.64$, $p < .15$. Moreover, the reliability estimates for each variable were similar whether or not Gender was in the design. Thus, Gender was not included in the reported analyses.

[5]It is acknowledged that these analyses ignore the possible effects of Gender (male subjects talking to females vs. female subjects talking to males) and Time (first, second, and third conversations). A complete specification of effects including these factors would be: $Peer_1(Subject_{21}(Gender_2) \times Time_3) \times Measures_6$. This analysis was conducted, and it provides tests for six additional effects: G, T, GT, GM, TM, GTM. Of these, all were nonsignificant at $p > .33$ except for the main effect of Gender, which was significant at $p < .03$. Females suggested a higher mean change in the men's behavior than the men suggested in the women's behavior. Given, (1) concerns about the replicability of this effect (one of six tests was significant), (2) the impossibility of interpreting this result since peer gender and subject gender were not crossed, and (3) given the

fact that the inclusion of these factors makes it impossible to test Subject effects and the Subjects x Measures interaction, which are the effects of primary interest in this study, without pooling certain effects into the error term, it was decided to ignore the possible Gender main effect. This decision should have little effect on the simpler model featured in this report since Gender did not interact with any of the other factors.

[6]The appropriate error term for these confidence intervals is what Winer referred to as $\underline{MS}_{w.cell}$. It is a weighted average of the P(S) and P(S)M error terms, with a df approximation due to Satterthwaite (Winer, 1971, p. 530). For these data, $\underline{MS}_{w.cell} = .74$, $\underline{df} = 475.0$.

APPENDIX B

INFORMED CONSENT FORM

The Pennsylvania State University

Title of Investigation: Acquaintanceship Study

Investigator(s): Michael G. Dow, M.S.

This is to certify that I, _____, hereby agree to participate as a volunteer in a scientific investigation as an authorized part of the education and research program of The Pennsylvania State University under the supervision of Dr. W. Edward Craighead.

The investigation and my part in the investigation have been defined and fully explained to me by _____, and I understand his/her explanation. A copy of the procedures of this investigation and the description of any risks and discomforts has been provided to me and has been discussed in detail with me.

I have been given an opportunity to ask whatever questions I may have had and all such questions and inquiries have been answered to my satisfaction.

I understand that I am free to deny any answers to specific items or questions in interviews or questionnaires.

I understand that any data or answers to questions will remain confidential with regard to my identity.

I understand that, in the event of physical injury resulting from this investigation, neither financial compensation nor free medical treatment is provided for such a physical injury, and that further information on this policy is available from the Vice President for Research and Graduate Studies, 207 Old Main (865-6331).

I certify that to the best of my knowledge and belief, I have no physical or mental illness or weakness that would increase the risk to me of participation in this investigation.

I FURTHER UNDERSTAND THAT I AM FREE TO WITHDRAW MY CONSENT AND TERMINATE MY PARTICIPATION AT ANY TIME.

Date	Date of Birth	Subject's Signature

I, the undersigned, have defined and fully explained the investigation to the above subject.

Date	Investigator's Signature

Purpose of Study: The purpose of this study is to learn about the acquaintanceship process when two people meet each other for the first time.

Procedure: During the assessment session you will complete questionnaires on social behavior and engage in social conversation. Your individual responses to these questionnaires will be available only to the research staff on this project, and they will be stored in file cabinets in Dr. Craighead's locked laboratory. There should be no additional discomfort with the procedure, except that which is experienced when talking to someone for the first time.

Potential Benefits: You will receive exposure to the process of academic research in psychology and gain extra course credit in introductory psychology. The study should help psychologists better understand the acquaintanceship process.

Period of Time Required: Less than 1 hour.

Contact Person: You may contact Michael Dow, M.S. (865-1580) if you have any concerns, problems, or questions about the study.

APPENDIX C

ACQUAINTANCESHIP STUDY

We would like you to meet and talk for 10 minutes with a student of the opposite sex. We are interested in learning about the acquaintanceship process when people meet each other for the first time. Please talk with the other person in the way you would typically get acquainted with someone. The other person is a student, and is also participating in this project for extra course credit. In order to make this a typical conversation, we ask that you do not talk about your participation in this project, or comment on the fact that this is a psychology experiment. Please talk with each other in the way you would normally get acquainted with someone.

The conversation will be observed by the assistant through a one-way window. She will knock on the door and enter the room when 10 minutes is up. After you finish the conversation, you will fill out a short form.

Instructions: Please answer each question by circling the appropriate number. In each instance, compare yourself with how other students your age and sex would seem in this same conversational setting.

1. How socially skilled do you expect to seem to the other person?

1	2	3	4	5	6	7	8	9
Much less than most people				Average				Much more than most people

2. How anxious or nervous do you expect to seem to the other person?

1	2	3	4	5	6	7	8	9
Much less than most people				Average				Much more than most people

3. How much do you expect the other person to like you?

1	2	3	4	5	6	7	8	9
Much less than most people				Average				Much more than most people

4. How physically attractive do you expect to appear to the other person?

1	2	3	4	5	6	7	8	9
Much less than most people				Average				Much more than most people

APPENDIX E

POST-CONVERSATION RATINGS

First name of person you talked with? _____

Male_____ or female_____?

Did you already know this person? _____ If yes, please explain how well you knew them. _____

Instructions: Please answer each question by circling the appropriate number. In each instance, compare yourself with how other students your age and sex would have seemed in this same conversational setting.

1. How socially skilled did you seem to the other person?

1	2	3	4	5	6	7	8	9
Much less than most people				Average				Much more than most people

2. How anxious or nervous did you seem to the other person?

1	2	3	4	5	6	7	8	9
Much less than most people				Average				Much more than most people

3. How much were you liked by the other person?

1	2	3	4	5	6	7	8	9
Much less than most people				Average				Much more than most people

4. How physically attractive did you seem to the other person?

1	2	3	4	5	6	7	8	9
Much less than most people				Average				Much more than most people

APPENDIX F

PARTNER RATINGS

Instructions: Please answer each question by circling the appropriate number. In each instance, rate the person you just spoke with, compared with how other students the same age and sex would seem in this same conversational setting. These ratings will not be shown to the other person. Please do not discuss them with anybody.

1. How socially skilled was the other person?

1	2	3	4	5	6	7	8	9
Much less than most people				Average				Much more than most people

2. How anxious or nervous was the other person?

1	2	3	4	5	6	7	8	9
Much less than most people				Average				Much more than most people

3. How much did you like the other person?

1	2	3	4	5	6	7	8	9
Much less than most people				Average				Much more than most people

4. How physically attractive was the other person?

1	2	3	4	5	6	7	8	9
Much less than most people				Average				Much more than most people

APPENDIX G

SUGGESTED CHANGES IN PARTNER BEHAVIOR

1. What changes (if any) in the other person's conversational style or behavior would make them seem more socially skilled? Please try to list specific types of behavior instead of general things, and list only those things that are possible to change.

2. What changes (if any) could the other person make to become more physically attractive? Please try to list specific things, and only list things that are possible to change.

Specific Behavioral Ratings

Part 1--Suggested Changes in Partner Behavior

Instructions: Please indicate what changes in the other person's conversational behavior (if any) would make them seem more socially skilled.

1. Verbal attention directed toward the person they talk with (e.g., ask questions and talk about the other)

1	2	3	4	5	6	7	8	9
Should have much less			Don't change					Should have much more

2. Verbal stammers, body trembling, fidgets

1	2	3	4	5	6	7	8	9
Should have much less			Don't change					Should have much more

3. Amount of time they talk

1	2	3	4	5	6	7	8	9
Should have much less			Don't change					Should have much more

4. Eye contact

1	2	3	4	5	6	7	8	9
Should have much less			Don't change					Should have much more

5. Hand gestures

1	2	3	4	5	6	7	8	9
Should have much less			Don't change					Should have much more

6. Pauses in the conversation, or pauses before they speak

1	2	3	4	5	6	7	8	9
Should have much less			Don't change					Should have much more

7. Foul language or swearing

1	2	3	4	5	6	7	8	9
Should have much less				Don't change				Should have much more

8. Extensiveness of vocabulary--use of "big words"

1	2	3	4	5	6	7	8	9
Should have much less				Don't change				Should have much more

9. Smiling

1	2	3	4	5	6	7	8	9
Should have much less				Don't change				Should have much more

10. Compliments to the person they talk with

1	2	3	4	5	6	7	8	9
Should have much less				Don't change				Should have much more

11. Self-praise and other positive statements about themself

1	2	3	4	5	6	7	8	9
Should have much less				Don't change				Should have much more

12. Criticism of the other person they talk with

1	2	3	4	5	6	7	8	9
Should have much less				Don't change				Should have much more

13. Positive statements about other people (third persons) or things

1	2	3	4	5	6	7	8	9
Should have much less				Don't change				Should have much more

Part 2--Impact of Specific Changes

Instructions: For each change you suggested in the last section (if there were any), rate how much improvement there would be in the other person's social skill if they were to change in the way you suggested. If you did not suggest a change before, then circle 0 to indicate no change in this section also. Please refer back to your ratings in the previous section as you consider each of these items.

1. Verbal attention directed toward the person they talk with (e.g., ask questions and talk about the other)

0	1	2	3	4	5	6	7	8
No change				Moderate improvement				Tremendous improvement

2. Verbal stammers, body trembling, or fidgets

0	1	2	3	4	5	6	7	8
No change				Moderate improvement				Tremendous improvement

3. Amount of time they talk

0	1	2	3	4	5	6	7	8
No change				Moderate improvement				Tremendous improvement

4. Eye contact

0	1	2	3	4	5	6	7	8
No change				Moderate improvement				Tremendous improvement

5. Hand gestures

0	1	2	3	4	5	6	7	8
No change				Moderate improvement				Tremendous improvement

6. Pauses in the conversation, or pauses before they speak

0	1	2	3	4	5	6	7	8
No change				Moderate improvement				Tremendous improvement

7. Foul language or swearing

0	1	2	3	4	5	6	7	8
No change				Moderate improvement				Tremendous improvement

8. Extensiveness of vocabulary--use of "big words"

0	1	2	3	4	5	6	7	8
No change				Moderate improvement				Tremendous improvement

9. Smiling

0	1	2	3	4	5	6	7	8
No change				Moderate improvement				Tremendous improvement

10. Compliments to the person they talk with

0	1	2	3	4	5	6	7	8
No change				Moderate improvement				Tremendous improvement

11. Self-praise and other positive statements about themself

0	1	2	3	4	5	6	7	8
No change				Moderate improvement				Tremendous improvement

12. Criticism of the person they talk with

0	1	2	3	4	5	6	7	8
No change				Moderate improvement				Tremendous improvement

13. Positive statements about other people (third persons) or things

0	1	2	3	4	5	6	7	8
No change				Moderate improvement				Tremendous improvement

VITA

Michael G. Dow
Date and place of birth: February 9, 1956; Eugene, Oregon.

Education

B.S., Psychology, and B.S., General Science, University of Oregon, 1979.
M.S., Psychology, The Pennsylvania State University, 1981.
Ph.D., Psychology, The Pennsylvania State University, 1983.

Academic Awards

Phi Beta Kappa, 1978
National Institute of Mental Health Fellowship, 1979-1980
Graduate School Fellowship, 1980-1981
Graduate School Fellowship, 1982-1983

Professional Experience

Instructor or teaching assistant for courses in behavior modification and statistics, The Pennsylvania State University, 1981-1982.
Therapist, Psychological Clinic and Center for Counseling and Adult Psychological Services, The Pennsylvania State University, 1979-1982.
Statistical analysis consultant, independent contractor, 1978-1983.

Publications

Bricker, D. D., & Dow, M. G. Early intervention with the young severely handicapped child. Journal of the Association for the Severely Handicapped, 1980, 5, 130-142.

Biglan, A., & Dow, M. G. Toward a second-generation model: A problem specific approach. In L. Rehm (Ed.), Behavior therapy for depression: Present status and future directions. New York: Academic Press, 1981.

Dow, M. G., Glaser, S. R., & Biglan, A. The relevance of specific conversational behaviors to ratings of social skill: An experimental analysis. Journal of Behavioral Assessment, 1981, 3, 233-242.

Dow, M. G., Craighead, W. E., & Borkovec, T. D. Instructional demand and pretesting in the assessment of social anxiety. Behavior Therapy, 1983, 14, 326-332.

Craighead, W. E., Kennedy, R. E., Raczynski, J. M., & Dow, M. G. Affective disorders-unipolar. In S. M. Turner & M. Hersen (Eds.), Adult psychopathology: A behavioral perspective. New York: Wiley, in press.

Dow, M. G., & Craighead, W. E. Cognition and social inadequacy: Relevance in clinical populations. In P. Trower (Ed.), Cognitive approaches to social skills training. Oxford: Pergamon, in press.