

Reliability of Heterosocial Skills Measurement with Sex Offenders

Barrie Byrnes Alexander¹ and Suzanne Bennett Johnson¹

Accepted May 15, 1980

The heterosocial skills of 79 sex offenders were assessed using two self-report inventories, the Social Anxiety and Distress Scale (SAD) and the S-R Inventory of Anxiousness (SRI). In addition, staff members rated the offenders on modified versions of the SAD and the Heterosocial Skills Behavior Checklist for Males (HSB). Twenty-one offenders were also videotaped during a brief interaction with a female confederate. The tapes were rated using the HSB. All ratings and videotapings were conducted twice within a two-week period. The two self-report scales showed good test-retest reliability ($r = .88$) and correlated $r = .63$ with each other. However, the SAD correlated significantly with the Crowne-Marlowe Social Desirability Scale and showed significant change from first to second administrations. The SRI showed less reactivity to social desirability and changed less over time. The staff ratings showed higher test-retest ($r = .72-.76$) than interrater ($r = .36-.43$) reliability. They correlated $r = -.56$ with each other and did not show a significant relationship to the Crowne-Marlowe. However, both staff-ratings changed significantly from test to retest. Inspection of correlations between staff and self-report measures indicated that these instruments may be strongly affected by rater variance. Different measures completed by the same person correlated higher than the same measure completed by different persons. The HSB showed high levels of interrater agreement (90–100%) and split-half reliability ($r = .95-.97$). However, test-retest reliability was poor for the overall score and for the category measuring form of conversation. Categories measuring voice and affect were moderately reliable over time ($r = .53-.74$). However, the voice category scores changed substantially from first to second videotaping. On almost all measures used in this study, the sex offenders' scores were similar to those derived from normal populations. Only the category measuring affect on the HSB seemed to differentiate this group from normals.

KEY WORDS: heterosocial skills measurement; sex offenders.

¹ Departments of Clinical Psychology and Psychiatry, University of Florida, Gainesville, FL 32610.

INTRODUCTION

The psychological literature on sex offenders consists largely of clinical descriptions in which the offenders' poor heterosocial skills are often emphasized. Both Gebhard et al. (1965) and Cohen et al. (1969) have reported that inadequate social functioning is characteristic of at least some pedophile populations. In this country, such presumed deficits are frequently treated by social skills training (Abel et al., 1976; Burkhardt, 1974; Laws & Serber, 1975; Stockton et al., 1977). Despite this active interest in the heterosocial characteristics of sex offenders, there is little empirical evidence addressing heterosexual skills measurement in this population.

Assessment devices designed for heterosocial measurement have been developed primarily with the student "minimal dater" population. Since no one modality is sufficient to decipher a behavior as complex as social skill, self-report instruments are generally used in conjunction with behavioral measures (Eisler, 1976; Hersen et al., 1973; Hersen and Bellack, 1977). A few studies have evaluated the reliability and validity of these measures (Arkowitz et al., 1975; Borkoveck et al., 1974; Rehm & Marston, 1968; Twentyman & McFall, 1975). However, Hersen and Bellack (1977) conclude in their recent review that "no single device has yet been clearly and thoroughly validated. Reliability (stability and homogeneity) has been an especially neglected issue" (p. 549). Furthermore, their utility as measures of heterosocial skill with sex offenders is unknown.

The present study assessed the reliability of two self-report instruments, two staff-rating scales, and one behavioral sampling procedure as measures of heterosocial skills with sex offenders. The intercorrelations between instruments were also examined. For the self-report and behavioral measures, instruments that showed the most promise were selected based on the available literature. These included the Social Anxiety and Distress Scale (Watson & Friend, 1969), the S-R Inventory of Anxiousness (Endler et al., 1962) and the Heterosocial Skills Behavior Checklist for Males (Barlow et al., 1977). The staff rating scales were developed for this study by modifying the Social Anxiety and Distress Scale and the Heterosocial Skills Behavior Checklist for Males. A repeated measures design was employed in which all measures were given twice within a two-week period in order to assess test-retest reliability.

METHOD

Subjects

All sex offenders from all levels of treatment at the North Florida Evaluation and Treatment Center, Gainesville, Florida, were asked to participate in the study. Very few refused, resulting in a sample of 79 men. These were involuntary,

court committed offenders. Thirty-nine percent of these were rapists or men who attempted rape, while 40 percent were pedophiles (evenly divided into homosexual and heterosexual), another 15% had had incestuous relations with their own children, and 7% fell into other categories (e.g., exhibitionists, voyeurs). Of these, 66% were under age 31, 22% were aged 31–45, and 12% were over 45. Forty-six percent of the sample had never been married, while 54% were married or divorced. Sixteen percent were black, the remainder white.

All 79 subjects were given self-report inventories and were rated by two staff members. Twenty-one subjects were chosen from the white population of recent intakes for the behavioral assessment. These were randomly selected and generally mimicked the whole population with respect to crime (rapist, pedophile, other), age (over 31 or under 31) and marital status (never married or married/divorced). Only white subjects were used, because Barlow et al. (1977) reported some differences between blacks' and whites' responses on the Heterosocial Skills Behavior Checklist for Males.

Staff

Eight experienced staff members who were in daily contact with the men provided the staff ratings. These were the primary therapists and caretakers in the residential buildings and were under the supervision of professional level staff. Staff members rated only those men whom they knew well. Each subject was independently rated by two staff members.

Instruments

Self-Report

The Social Anxiety and Distress Scale (SAD) is a 28-item true-false scale designed to measure distress, discomfort, fear, and anxiety in social situations. The authors report moderate test-retest reliability ($r = .68$) with a KR 20 index of homogeneity of .94 (Watson and Friend, 1969). Correlations with the Crowne-Marlowe Scale of Social Desirability (Crowne and Marlowe, 1960) were low and negative (Watson and Friend, 1969). Arkowitz et al. (1975) found significant differences on the SAD between low and high frequency daters in a college population. Two studies reported significant decreases in SAD scores following treatment (Christensen & Arkowitz, 1974; Curran et al., 1976). However, in a study by McGovern et al. (1975), the SAD seemed to be insensitive to treatment effects.

The S-R Inventory of Anxiousness (SRI) asks the subject to rate on a five-point scale the degree to which he would experience each of 14 different responses (e.g., heart beats faster, perspires, feels exhilarated) in each of five

hypothetical situations. The situations developed for use in the present experiment were as follows:

1. You go to the beach or pool and there is a nice looking woman there you would like to know better.
2. You are calling a woman you like a lot to ask her to do something with you.
3. You are in line at the grocery store and a woman turns to you, commenting about the high price of food.
4. You are trying to make casual conversation with the woman in line in front of you at the bank.
5. You are introduced to an attractive girl and are left to make conversation with her for a few minutes.

Arkowitz et al. (1975) did find significant differences between the SRI scores of high and low frequency daters, which gives the inventory some validation support. No reliability or standardization data are available. Christensen and Arkowitz (1974) reported a significant decrease in SRI scores post-treatment, while others have not found such changes (MacDonald et al., 1975; McGovern et al., 1975).

Because research indicates that self-report of males tends to be less accurate than that of females and may reflect an unwillingness for men to report anxiety (Futch and Lisman, 1977; Gambrill & Ritchey, 1975; Mischel, 1970), the Crowne-Marlowe Social Desirability Scale (Crowne & Marlowe, 1960) was also administered. This is a 30-item true-false scale that measures the subject's tendency to answer in a socially desirable direction.

Behavioral Assessment

The Heterosocial Skills Behavior Checklist for Males (HSB) was used to rate the behavior of 21 subjects during a brief videotaped interaction with a female confederate. The procedures described by Barlow et al. (1977) were generally followed. Four confederates (two college age, two in their late 30's) were trained to refrain from initiating conversation or asking questions and to limit their responses to about five words. They were also taught to inhibit frequent nonverbal communication such as smiling, yet not be unfriendly. A confederate of the same race and age as the subject was seated in the videotaping room when the subject arrived. Before entering the videotaping room the subject was told:

For the next 5 minutes you will be conversing with a woman while being videotaped. Some people find it helpful to imagine that they are in a setting like a bus station or a waiting room. I'd like you to do whatever you would normally do if this was someone you found attractive and would like to know better.

After inquiring whether he had any questions, the subject was then escorted into the videotaping room and introduced to the confederate, after which the experimenter left. If the subject initiated no conversation after approximately 2 minutes, the confederate commented on the weather: "It certainly is a (rainy, sunny) day today." If the subject did not respond to that overture, the confederate did not initiate any more conversation and the videotaping proceeded until the full 5-min time period elapsed. At that point, the experimenter interrupted and thanked the subject.

Each videotape was then rated by trained observers using the HSB. The original version included items in four categories: voice, form of conversation, affect, and motor behavior. Barlow et al. (1977) reported that scores in these categories showed high interrater reliability. Their data also suggested that scores in three of the categories, voice, form of conversation, and affect, could successfully differentiate socially inadequate from socially adequate males. The motor behavior category was not used in the present study as it failed to differentiate between Barlow et al.'s socially adequate and inadequate groups.

Staff Ratings

A modified form of the SAD (SAD-staff) containing the same items but worded so that the staff could answer, was used (e.g., instead of "I feel. . ." it read "he feels. . ."). Staff were instructed to answer the items in the way they felt was the most accurate for the offender. In addition to "true or false" they had the option of "don't know."

The same staff members were also asked to do a behavioral rating using the categories rated on the HSB (HSB-staff). They were instructed to rate the subject's behavior as appropriate or inappropriate for each category, based on their observations of the subject's interactions with female staff.

Procedure

All 79 subjects completed the self-report measures twice and were rated twice by two staff members at a two-week interval. The 21 subjects selected for the videotaped interaction with a female confederate were also videotaped twice two weeks apart.

Half of the subjects videotaped were taped before answering any self-report measures. Half were assessed after completing the initial self-report instruments. All self-report and staff-rating devices had two forms. One form was as published, while the other had identical items in random order. The order of

each form's presentation was also randomized. To increase cooperation, both subjects and staff were informed that they would be asked to repeat the assessment procedures.

Scoring

Total scores were computed for each of the self-report inventories (SAD, SRI, Crowne-Marlowe) and staff ratings (SAD-staff, HSB-staff) at test and retest.

The two videotapes of each of the 21 subjects' interactions were divided into 30-sec blocks for rating purposes and were independently observed by two raters. The raters were trained in discriminating responses defined on the HSB code until they were at least 80% reliable. Each rater then coded five tapes each session. One of the five tapes was identical for both raters and was used for checking reliability. Raters were aware that their reliability was being checked, but were unaware of which tape was being used for this purpose.

Scoring procedures followed those of Barlow et al. (1977). Scores were obtained for each category in addition to an overall score that combined all three categories. Agreement between raters was calculated by noting agreements on appropriate and inappropriate behaviors as well as nonoccurrence of behavior during each 30-sec block for each category. Agreements were divided by agreements plus disagreements to arrive at a reliability coefficient. Reliability coefficients were calculated for each of the three categories as well as for the overall score.

RESULTS

Self-Report and Staff Ratings

The SRI correlated $r = .88$ ($p < .0001$) with itself, and the SAD also had a test-retest correlation of $r = .88$ ($p < .0001$). The SAD-staff had a correlation of $r = .72$ ($p < .0001$) while the HSB-staff showed a test-retest correlation of $r = .76$ ($p < .0001$).

For the staff measures, it was possible to calculate interrater reliability since two staff rated each subject. On the SAD-staff, this correlation was $r = .431$, ($p < .0001$). It was $r = .357$ for the HSB-staff ($p < .0012$).

The correlations between the two self-report measures of social skill and the two staff measures are presented in Table I. The two self-ratings correlate highly with each other ($r = .63$) as do the two staff ratings ($r = -.56$). Each of these correlations are lower than the test-retest correlations of each test with itself. Note that the correlations in which the same rater completed different

Table 1. Correlation Coefficients of Self-Report and Staff Measures

	Self SAD	Staff	
		SAD-staff	HSB-staff
Self			
SRI	$r = .63$ $p < .001$	$r = .27$ $p < .02$	$r = -.25$ $p < .03$
SAD		$r = .35$ $p < .001$	$r = -.24$ $p < .03$
Staff			
SAD-staff			$r = -.56$ $p < .001$

Note: SAD = Social Anxiety and Distress Scale; SRI = S-R Inventory of Anxiousness; HSB-staff = modified version of the Heterosocial Skills Behavior Checklist for Males.

instruments (correlations between HSB-staff and SAD-staff and between SAD and SRI) are higher ($r = -.56$ and $.63$, respectively) than that in which the same instrument (SAD, SAD-staff) was completed by different raters ($r = .35$). However, this latter correlation is higher than those between different instruments completed by different raters.

Scores obtained at the first and second test administration were compared using the F statistic. Statistically significant differences were found for all measures except the SRI. Differences on the SRI approached significance: SAD, $F(1,78) = 15.24$, $p < .0002$; SAD staff, $F(1,78) = 6.41$, $p < .031$; HSB staff, $F(1,78) = 9.78$, $p < .0025$; SRI, $F(1,78) = 3.42$, $p < .07$. The SAD, SRI, and HSB staff measures changed in the direction of decreased anxiety and more appropriate behavior, while the SAD-staff scores reflected an increase in anxiety at retest.

The SAD staff and HSB staff both showed low and nonsignificant correlations with the Crowne-Marlowe Social Desirability Scale. The SAD and SRI, however, showed slightly higher statistically significant correlations, $r = -.42$, $p < .0001$, and $r = -.25$, $p < .02$, with this instrument.

Analyses of variance done to assess the effects of age (under 31 or over 31), marital status (never married or married/divorced) and crime (rape/attempted rape, pedophile, or other) on the self-report and staff-rating scores revealed a significant effect for age on the SAD, $F(1,67) = 10.91$, $p < .0015$, and the SAD-staff, $F(1,67) = 5.21$, $p < .026$. On the SAD, the younger subjects (under 31) scored in the direction of being more anxious than the older subjects. Similarly, on the SAD-staff they were perceived as being more anxious by the staff.

Behavioral Ratings

For the HSB, interrater, split-half, and test-retest reliability were assessed. During training, overall interrater reliability reached 93% agreement, and although raters were told they would have to retrain if they fell below 80% agreement, this was never necessary. During actual rating sessions, interrater reliability overall and for each of the three categories was consistently between 90 and 100%.

Split-half reliability was calculated for voice, form, and affect by comparing scores obtained from the odd-numbered 30-sec segments to those obtained from the even-numbered segments for the first videotapes. Correlations were high for each category and the overall score, ranging from $r = .95$ to $.97$.

The data for test-retest reliability on the HSB was initially analyzed by percentage as Barlow et al. (1977) did, but a decision was made to analyze the raw data, because ceiling effects (on voice) made a percentage correlation inappropriate. Test-retest reliability for the overall score on this measure was poor, $r = .27$, $p < .25$. However, voice, $r = .74$, $p < .001$, and affect, $r = .53$, $p < .02$, were moderately reliable, while form, $r = .09$, $p < .72$, was unreliable over time. The test-retest reliability of an overall score combining voice and affect with form deleted was $r = .45$, $p < .04$.

Whether the scores changed significantly from test to retest was also assessed using the F statistic. Only voice showed a tendency to change over time, $F(1,19) = 4.08$, $p < .06$; all subjects improved and most received the maximum appropriate score on this measure at retest.

To assess whether being videotaped had an effect on self-report, a t -test was done comparing those subjects who completed the self-report inventories before videotaping to those who completed them after. While there was no significant effect on the subjects' tendency to answer in a socially desirable direction on the Crowne-Marlowe, there was a significant effect on both the SAD, $t(19) = -3.1$, $p < .006$, and the SRI, $t(16) = -2.86$, $p < .01$. Those who completed the social skills inventories after being videotaped scored in the direction of being more anxious than those who completed the inventories before videotaping.

Comparison with Normals

Since the treatment of sex offenders for their poor social skills is predicated on the assumption that they are lacking those skills, a comparison of their scores with those of normals is of some interest.

When this study's sex offenders' scores on the Crowne-Marlowe were compared to those of the 120 college students used to standardize the instrument (Crowne and Marlowe, 1960), the means were within one standard deviation of

Table II. Mean Scores from Present Study and Standardization Studies on Two Self-Report Measures

Measures	<i>N</i>	Sex offenders	<i>N</i>	College students
CM	79	$\bar{X} = 14.77$ (SD = 7.07)	120	$\bar{X} = 13.72$ (SD = 5.78)
SAD	79	$\bar{X} = 10.86$ (SD = 7.8)	297	$\bar{X} = 9.11$ (SD = 8.01)

Note: CM = Crowne-Marlowe Social Desirability Scale; SAD = Social Anxiety and Distress Scale. Data for college students was taken from Crowne and Marlowe (1960) and Watson and Friend (1969).

each other (see Table II). The sex offenders' scores fell in a generally normal distribution with 61% within one standard deviation of the mean.

The same was true of the SAD scores when compared with those of the undergraduates tested by Watson and Friend (1969) in their standardization study (see Table II). Again the sex offenders' scores were generally normally distributed with 61% within one standard deviation of the mean.

Unfortunately, a similar comparison for the SRI was not possible since there is no standardization data for this measure and each investigator tends to use different stimuli with the standard responses developed by Endler et al. (1962). The version of the SRI used in the present study had a mean of 139.24 with a standard deviation of 38.93 at initial testing. At retest, the mean was 135.32 (SD = 37.29).

In developing the HSB, Barlow et al. (1977) used sex offenders judged highly socially inappropriate and high school students rated highly appropriate. A comparison of Barlow et al.'s data with that from the present study is presented in Table III. As a group, the sex offenders in the present study consistently scored between Barlow's two groups. For voice and form, however, they scored more like the highly appropriate high school students, while on affect their scores were closer to those of the inappropriate sex offenders. These data

Table III. Mean Percentage Scores from Present Study and Standardization Study on Heterosocial Skills Behavior Checklist

Measure	Appropriate high school students (10)	Inappropriate sex offenders (10)	Sex offenders (21)	
			Time 1	Time 2
Voice	$\bar{X} = 96\%$	$\bar{X} = 84\%$	$\bar{X} = 89\%$	$\bar{X} = 99.6\%$
	SD = 7.4	SD = 16.6	SD = 5.6	SD = .56
Form	$\bar{X} = 95\%$	$\bar{X} = 67\%$	$\bar{X} = 86\%$	$\bar{X} = 90\%$
	SD = 10.8	SD = 26.9	SD = 7.8	SD = 3.07
Affect	$\bar{X} = 93\%$	$\bar{X} = 61\%$	$\bar{X} = 73\%$	$\bar{X} = 74\%$
	SD = 10.1	SD = 22.3	SD = 3.5	SD = 2.7

Note: Data for 10 appropriate high school students and 10 inappropriate sex offenders are from Barlow et al. (1977). Remaining data are from present study.

are not surprising in view of the fact that Barlow et al. selected groups extremely different in heterosocial ability, while the present study focused on a general population of involuntary, court-committed sex offenders.

DISCUSSION

The social skills of some sex offenders are thought to be so inadequate as to be contributory to their crimes. Despite this, few attempts have been made to measure social skills of these offenders, and measurement devices have not been assessed for use with this particular population. This study was an attempt to discover which of several social skills assessment instruments developed primarily with college students might be useful in this regard.

The two self-report scales selected, the SAD and SRI, showed good test-retest reliability. Both had high test-retest correlations and correlated moderately with each other, providing some evidence for convergent validity. However, these scales need to be compared to a third instrument designed to measure something other than social skills, in order to assess both convergent and discriminant validity.

When used with sex offenders, the SAD correlated more with the Crowne-Marlowe ($r = -.42$) than it did when used with the normal population described by Watson and Friend (1969) ($r = -.25$). Further, there were significant changes from test to retest, suggesting that prior experience with the SAD may influence subsequent responses to it.

The SRI showed less reactivity to social desirability ($r = .25$) and changed less over time. It does lack standardization data, however, and would benefit from the adoption of one form to be used by all investigators.

Staff rating scales such as those used in this study are nonexistent in the literature. For this reason it was necessary to modify the SAD and HSB for staff use. Both of these modified measures proved to be reliable over time with correlations ranging from .72 to .76. The data suggest some convergent validity, as the scales correlate moderately with each other ($r = -.56$) and had low but significant correlations with the two self-rating scales. Both also showed moderate interrater reliability. Unfortunately, each changed significantly from the first to the second administration, but in different directions. While it is possible that subjects were more anxious (SAD-staff) while demonstrating more appropriate social behavior (HSB-staff) at retest, this difference is difficult to interpret.

These staff rating scales had low negative correlations with the social desirability scale. This suggests that the staff is less suspect to any attempt to present the offenders in a "good" light than are the offenders themselves.

With the SAD it was possible to compare scores on the measure when used as a self-report instrument to scores when rated by the staff. Although the self- and staff-ratings correlated significantly to suggest convergent validity, the scores

on the SAD were very sensitive to who was doing the rating. One might expect the SAD to correlate more highly with the SAD-staff than with either the SRI or the HSB-staff. In fact, the SAD correlated higher with the SRI, $r = .63, p < .0001$, than it did with the staff version of the SAD, $r = .35, p < .0001$. Similarly, the SAD-staff showed higher correlations with the HSB-staff, $r = -.56, p < .0001$, than it did with the SAD, $r = .35, p < .0001$. Thus the staff ratings correlated with staff ratings, self-ratings with self-ratings, despite use of the same instrument by both staff and self. Similar comparisons are not possible for the SRI.

In this study, younger men reported themselves as more anxious (on the SAD) and were perceived as more anxious by the staff (on the SAD-staff). It is unclear whether this is a normal consequence of age and experience or whether it is a finding specific to sex offenders. For example, the younger offenders might be less criminally experienced, may be more likely to have been caught in a first offense, and therefore may be more anxious than the older man, who is an "old hand." Further research is needed to determine whether this age differential is specific to social anxiety and whether it is replicated in a normal population.

The behavioral test, the HSB, was the most costly in terms of time and manpower, because it involved videotaping and subsequent rating of the tapes. The high interrater reliability found by Barlow et al. (1977) was replicated. In the present study the raters knew they would be checked for reliability. They trained, without problem, to a high level of agreement and were able to easily maintain this level. Split-half reliability on this instrument was also very good. The overall test-retest reliability was not satisfactory, however. It may be possible to use just voice and affect for rating, but voice changed substantially over time and showed a ceiling effect for most subjects at retest. In addition, caution should be exercised in deciding when to administer other measures that may be affected by the videotaping, as were the SRI and SAD in this study.

During videotaping, few of the subjects chose to imagine that they were in another setting. This may indicate that this population tends to be more comfortable with concrete instructions. They were, however, very cooperative, and the chance to talk with a woman, even in such a structured setting, proved to be a highly desirable experience and also a challenge. They viewed their performance very seriously and some, who felt they had done poorly the first time, reported practicing between sessions. This may have contributed to the improvement in voice scores over time. It is interesting, however, that affect (consisting of measures of facial expression, eye contact, and laughter) did not change over time and was the only measure in this study that discriminated sex offenders as a group from normals. It is possible that most sex offenders function adequately in most interpersonal situations and only a small subgroup is in need of heterosocial skills training. It is also possible that the methods we are using to measure social skills are tapping aspects of social interactions that are not relevant to the treatment of most members of this population. Perhaps there are more subtle

behaviors that we are failing to measure. If this is true, the use of videotaping, although costly, is a necessity.

ACKNOWLEDGMENTS

The authors wish to thank Denise Edwards, Alicia Schmitt, Brian Rooke, and Irene Hernandez for their data collection efforts, and Dr. Randy Carter for his assistance in the data analysis. Also acknowledged are the residents and staff of the North Florida Evaluation and Treatment Center, Gainesville, Fla. Reprint requests should be sent to S. B. Johnson, Box J-234, Division of Child and Adolescent Psychiatry, Gainesville, FL 32610.

REFERENCES

- Abel, G. G., Blanchard, E. B., & Becker, J. V. Psychological treatment of rapists. In M. J. Walker & S. L. Brodsky (Eds.), *Sexual assault*. Lexington: Lexington Books, 1976.
- Arkowitz, H., Lichtenstein, B., 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. Heterosocial skills behavior checklist for males. *Behavior Therapy*, 1977, 8, 299.
- Borkoveck, T. D., Stone, N. M., O'Brien, G. T., & Kaloupek, D. G. Evaluation of a clinically relevant target behavior for analogue outcome research. *Behavior Therapy*, 1974, 5, 503-513.
- Burkhardt, P. L. *Orientation booklet for sexual reorientation*. Atascadero State Hospital, program IX, Dept. of Health, Atascadero, Calif., 1974.
- Christensen, A., & Arkowitz, H. Preliminary report on practice dating and feedback as treatment for college dating problems. *Journal of Counseling Psychology*, 1974, 21, 92-95.
- Cohen, M., Seghorn, T., & Calmas, W. Sociometric study of the sex offender. *Journal of Abnormal Psychology*, 1969, 74, 249-255.
- Crowne, D. P., & Marlowe, D. A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 1960, 24, 349-354.
- Curran, J. P., Gilbert, F. S., & Little, L. M. A comparison between behavioral replication training and sensitivity training approaches to heterosexual dating anxiety. *Journal of Counseling Psychology*, 1976, 23, 190-196.
- Eisler, R. M. Behavioral assessment of social skills. In M. Hersen & A. S. Bellack (Eds.), *Behavioral assessment: A practical handbook*, Oxford, U.K.: Pergamon, 1976.
- Endler, N. S., Hunt, J. M., & Rosenstein, A. J. An S-R inventory of anxiousness. *Psychological Monographs*, 1962, 76 (17 #536).
- Futch, E. J., and Lisman, S. A. Behavioral validation of an assertiveness scale: the incongruence of self-report and behavior. Paper presented at AABT, Atlanta, December, 1977.
- Gambrill, E. D., & Ritchey, C. A. An assertion inventory for use in assessment and research. *Behavior Therapy*, 1975, 6, 550-561.
- Gebhard, P. H., Gagnon, J. H., Pomeroy, W. P., & Christenson, C. V. *Sex offenders*. New York: Harper & Row, 1965.
- Hersen, M., & Bellack, A. S. Assessment of social skills, In A. R. Ciminero, K. S. Calhoun, & H. E. Adams (Eds.), *Behavioral Assessment*. New York: Wiley, 1977.
- Hersen, M., Eisler, R. M., & Miller, P. M. Development of assertive responses: clinical measurement and research considerations. *Behaviour Research and Therapy*, 1973, 11, 505-521.

- Laws, D. R., & Serber, M. Measurement and evaluation of assertive training with sexual offenders. In R. E. Hosford & C. S. Moss (Eds.), *The crumbling walls: treatment and counseling of prisoners*. Champaign, Ill.: University of Illinois Press, 1975.
- MacDonald, M. L., Lindquist, C. U., Kramer, J. A., McGrath, R. A., & Rhyne, L. L. Social skills training: the effects of behavior rehearsal in groups of dating skills. *Journal of Counseling Psychology*, 1975, 22, 224-230.
- McGovern, K. B., Arkowitz, H., & Gilmore, S. K. Evaluations of social skills training programs for college dating inhibitions. *Journal of Counseling Psychology*, 1975, 22, 505-512.
- Mischel, W. Sex-typing and socialization. In P. Mussen (Ed.), *Manual of child psychology* (Vol. 2). New York: Wiley, 1970.
- Rehm, L. P., & Marston, A. R. Reduction of social anxiety through modification of self-reinforcement: an instigation technique. *Journal of Consulting and Clinical Psychology*, 1968, 32, 565-574.
- Stockton, M. D., Legum, L., and McAnaney, M. A comprehensive treatment program for sex offenders. Unpublished manuscript, 1977 (available from the North Florida Evaluation and Treatment Center, Gainesville, Fla.).
- Twentyman, C. T., & McFall, R. M. Behavioral training of social skills in shy males. *Journal of Consulting and Clinical Psychology*, 1975, 43, 384-395.
- Watson, D., & Friend, R. Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology*, 1969, 33, 448-457.