Evaluating Interpersonal Skills Training for Public Welfare Staff

Elizabeth W. Lindsey
University of Georgia

Donald B. Yarbrough
University of Iowa

Thomas D. Morton
Child Welfare Institute

A week-long in-service training session focused on improving interpersonal helping skills for 116 public welfare staff. The program was evaluated to assess changes in participants' knowledge, skill mastery, and self-perceptions. Results indicate statistically significant gains on use of attending behaviors, reflections, and a global rating of empathy, respect, and genuineness. Discussion centers on the practical significance of the small gains noted and on the need for further research to establish threshold time periods and appropriate instructional strategies necessary for substantial skill gains.

Human service and counseling agencies rely increasingly on short-term training to equip staff with the skills and knowledge needed to resolve complex practice problems. Evaluation of the effectiveness of such training is critically needed. Without more knowledge about factors that enhance learning in short-term programs, effective programs cannot be purposefully designed.

While interpersonal helping skills programs in professional education have been studied extensively,[1] less attention has been paid to the

evaluation of in-service training programs for these skills.[2] For many human service workers and counselors, the first or only exposure to formal instruction in interpersonal helping skills comes through in-service training programs. Such programs vary greatly in both instructional strategies and length of training. Often the rationale for the length of a training program seems more related to the amount of time agency administrators or training designers deem available than to the amount of time needed for skill mastery.

Little is known about the threshold training time that is required to become proficient in certain interpersonal counseling skills. Mitchell found that while maximum effects in respect, concreteness, and genuineness were reached after 18 hours of training, empathy continued to improve through 54 hours.[3] However, mean empathy scores for 35 subjects were only 2.73 after 54 hours, still below the 3.0 considered to be the minimum effective level for helping, on the Carkhuff Scale used, raising the issue of how much improvement on skills is possible through short-term approaches. One- or 2-day communication skills workshops may be a waste of time and money. Other factors such as types of learning activities and opportunity for practice and feedback may also detract from or enhance the effectiveness of short-term training.[4] The lack of knowledge in this area suggests the need for a closer look at short-term interpersonal skills training programs, especially those designed as in-service or continuing education programs.

In 1980 the Georgia Division of Family and Children Services contracted the Office of Continuing Social Work Education (OCSWE) at the University of Georgia to design a 30-hour interpersonal skills training program called Casework Interviewing Effectiveness: Level I (CIE I). In addition, the division financed an evaluation of this program. The evaluation design had to meet two requirements. First, it had to reliably assess the effectiveness of the CIE I training. Second, measurement had to fit naturally into the training format and facilitate rather than hinder the training process.

The following sections describe how these requirements were met and the results of the evaluation. The results raised questions concerning the practical significance of skill improvements noted among participants. In the concluding section, the authors discuss some of the implications for designing and evaluating training programs based on this experience.

Program Description

Purpose

The training program was developed to train public welfare staff to use essential interviewing skills. The 30-hour program focused on the interpersonal skills that facilitate effective interviewing.

The goal of training was to assure more effective delivery of services to clients while enabling workers simultaneously to communicate caring, concern, and respect for clients as individuals. The program emphasized techniques for developing and maintaining a positive working relationship with clients while accomplishing the tasks of the interview (e.g., determination of eligibility for assistance, developing a contract, discussing child discipline techniques).

The CIE I training program was based on the microskills identified by Ivey and Authier:[5] attending behaviors, reflections, use of questions, summarization, self-disclosure, information giving, directions, and advice. In addition, three "conditions of helping" that have been posited as necessary for effective helping[6] were also taught: empathy, respect, and genuineness.

Ivey and Authier[7] have suggested that interpersonal skills are generic, and that skills used by practitioners who have been trained in various paradigms and who work in diverse helping settings do not differ substantially. Rather, the pattern of skill use, including the frequency of certain skills and the content focused upon by the helper, varies according to the nature and purpose of the interview. Morton and Lindsey have proposed a model for effective use of interpersonal helping skills in public welfare practice.[8] The CIE I training program, based on this model, was designed to teach generic skills to both income maintenance and social services staff. Both videotape demonstrations of skills and exercises incorporated examples from both types of public welfare interviews to assure the relevance of the training program to both groups of staff.

Training Format

CIE I was designed as a week-long program that provided an intensive training experience. Each skill was taught in a separate module. At various points during the week, participants engaged in exercises that enabled them to integrate their use of several skills. Each skill module followed a format in which participants gained knowledge about the skill, observed a model of the skill, had the opportunity to practice the skill, and received feedback on their performance. At the core of each module was an instructional videotape that provided content related to the purpose and appropriate use of each skill, examples of skill use in practice, and exercises. The skill examples were drawn primarily from actual interviews between public welfare workers and clients, videotaped in county agency offices. Excerpts from actual interviews enabled participants to see how skills could be used in client interactions analogous to their own. In addition to the videotape exercises, each module incorporated other exercises (i.e., paper and pencil, group discussion, and role play), which gave participants opportunities for practice.

To practice and receive feedback on several skills in combination, participants made three videotapes. The first was made after the introduction to the workshop but before any skills training had taken place. Participants were assigned worker and client roles to play in simulated interviews. Individual guided self-assessments of the tapes followed later that evening and served to orient the participants to their own particular learning needs. The second videotape was made at the end of the second training day. Each person chose a personal topic to discuss with a partner, who then attempted to use the skills learned thus far in the program (attending, reflections, empathy, respect, and genuineness) to facilitate the talkers' exploration of the topics they chose to discuss. These tapes were viewed and discussed in small groups of four to eight participants, led by a trainer who assisted participants in giving each other feedback. The third videotape was made at the end of the training. Like the first tape, the content of the third videotape was work related, as participants were given worker and client roles to play in a simulated interview. In the worker role, participants practiced all skills learned during the program.

Participants

The participants involved in the CIE I training program were income maintenance workers from county welfare offices. Although CIE I was originally designed to train both service and eligibility staff, changes in availability of training funds prevented inclusion of social services staff in the training.

Nine CIE I workshops were held with a total of 116 participants over a 5-month period. Attendance at each workshop varied from a low of eight participants to a high of 16 participants. The participants, who were predominantly female, had an average age of 34 years, had been to school an average total of 14.8 years, and had an average of 5.3 years experience in human services work.

Evaluation Strategy and Methodology

The effectiveness of CIE I was examined from several perspectives. Since the training emphasized skills development, the most important measures evaluated changes in skill level before and after training. Other measures assessed participants' knowledge gain, changes in perceived ability, and satisfaction with training.

Measures

The four measures used were: (1) ratings of behaviors in simulated interviews, (2) a knowledge test, (3) an interviewing effectiveness self-report, and (4) participant evaluations.

Behavioral ratings.—Data on participants' use of interviewing skills were gathered by videotaping simulated eligibility interviews as a part of the training process. Consent forms were signed by all participants who were willing to have their tapes used for evaluation purposes. Only two of the 116 of participants declined to do so.

The pretraining assessments were based on the videotaped interviews conducted on the first day of training, immediately after the initial introduction to the workshop but prior to any skill modules. The posttraining assessments were based on videotaped interviews conducted on the next to the last day of the workshop, after all the skills had been taught. The videotaped interviews were simulated worker-client interviews and were designed to be an integral part of the training (participants received group feedback on their skill use in these interviews) as well as providing evaluation data. (An intermediate videotaping on the second day of training was not used for evaluation purposes.)

Eight client roles were prepared for the participants to use in the simulations. Trainees alternately assumed client and worker roles, which were given to them a few minutes before their simulation was to begin. The only observer in the room was a camera operator. Each simulated interview lasted approximately 8 minutes. A typical set of instructions to the participant playing the worker was: "You are conducting a 60-day case review for Jane Doe, age 34, and her son Francis Doe, age 12, AFDC and Food Stamp recipient. Father's whereabouts unknown. Left home unexpectedly. Did not report to work. Case referred to Child Support Recovery Unit, but without much hope of recovery because client did not have husband's social security number. Client unable to find work. Previous experience as a grocery store clerk. Client has not worked since birth of first child." The participant playing the client typically received much more thorough directions, including a number of statements about how he or she felt. These feelings included love, mistrust, anger, joy, sadness, frustration, fear, and shock. While there were no specific instructions to express these emotions, the narrative that the "client" read emphasized the feelings he or she was to portray.

Four skill areas were assessed. It was hypothesized that posttraining tapes would show improvements in: (1) demonstrated empathy, respect, and genuineness; (2) attending behaviors; (3) use of reflections; and (4) use of questions.

The measure of empathy, respect, and genuineness chosen was the global scale developed by Gazda.[9] This instrument rates helper statements on a scale from 1.0 (not helpful; hurtful) to 4.0 (underlying feelings reflected), with 3.0 (surface feelings reflected) considered the minimum level of effective response. Two experienced raters were trained to an average interrater reliability of .86. In addition, the raters received identical numbers of pre- and posttapes and were not told

which were the pre- and which were the posttraining tapes. Working with both the videotapes and typewritten transcripts, they rated trainee statements.

Instruments were developed by project staff to measure attending behavior, use of reflections, and use of questions. The measure of attending was based on effective nonverbal behavior as described by Benjamin, Egan, and Ivey and Authier and included ratings of eye contact, open posture, and forward lean.[10] During separate analyses of the tapes, the interviews were checked at 15-second intervals for the presence of each of the nonverbal behaviors. Eye contact was defined as looking directly at the person without turning the eyes away. Open posture was present if the worker placed both feet on the floor with extremities uncrossed and open. Forward lean was indicated by leaning forward in the chair with the back not touching the chair. A social work graduate student, without knowledge of the project and trained by project staff to an interjudge agreement of .94, conducted the analysis of attending behavior.

The measure of reflections consisted of a frequency count of reflections of content, reflections of feeling, simultaneous reflection of both feeling and content, and experiential reflections. A project staff member with primary responsibility for development of the reflections module conducted this analysis from transcripts without knowing which were pre- and which were posttraining. The rater did not train any other raters, and thus an interjudge agreement index was not computed. However, self-checks on consistency of ratings indicated that the rater's subsequent classifications of already-classified transcripts agreed more than 92% of the time.

The evaluation of questioning behavior was conducted with an instrument that analyzed the frequency of "general" open questions, "why" questions, closed questions, and multiple questions. Two social work graduate students trained to an interrater reliability of .94 analyzed the transcripts without knowing which were pre- and which were posttraining. Both raters analyzed all transcripts.

Knowledge test.—A criterion-referenced knowledge test was developed to assess participants' entry and exit levels. To assess participants' knowledge before training, the knowledge test was randomly assigned to one-half of the participants, who received it by mail prior to training. In most instances, the same instrument was administered again on the next-to-last training day. In other instances, however, it was mailed out after training. In all cases, participants were asked not to use any reference books and to answer the questions based on what they knew. All participants were advised that their test scores would be kept confidential, and that the sole use of the instrument was to evaluate and improve the training program.

Interviewing effectiveness self-report.—This measure was a 15-item Likert-type scale that assessed participants' perceptions of their own ability

prior to and after training. A total score was computed by summing ratings for all 15 items.

Participant evaluation.—Module-by-module evaluations were completed by participants to give specific feedback on individual modules and trainers. The data were used mainly to monitor group satisfaction with training during the week and to make changes when necessary. Participants also completed a summary evaluation of each week of training.

Design

The constraints on this training made it impossible to select participants randomly. Participants volunteered or were selected by their supervisors. Since the measurement procedures were to a large extent embedded in the training experience, it was not financially feasible or experimentally justifiable to select a group of potential participants for comparison measurement. Thus, the participants had to serve as their own controls. For the knowledge tests and the videotape assessments, different randomly selected groups were assigned to either a posttest-only condition or a pretest-posttest condition. This permitted the investigation of any unwanted pretest-posttest or pretest-treatment interactions. Because the data were collected during nine consecutive 1-week-long training sessions, "maturation" and/or "history"[11] were not credible alternative explanations for any pretest-posttest differences. The major concern regarding the results of these studies is that the participants did not represent a random sample of practitioners who might in the future receive this training.

In order to deal with problems inherent in self-reports as measures of training effectiveness, a different design was used for the interviewing effectiveness self-reports. Howard warns that pre- and posttraining reports of ability are often inaccurate because of a change in participants' ideas of what the ability is as a result of training.[12] People tend to rate themselves higher prior to training because of their ignorance of the true nature of the ability. Howard suggests the use of a retrospective self-report administered after training on which participants simultaneously rate their ability prior to training ("THEN") and after training ("POST"). For this study, it was not possible to collect data for all possible validity checks: only two randomly assigned groups were compared. Group 1 served as a pretest reactivity check by receiving only a posttraining POST self-report. Group 2 completed a pretraining self-report (PRE) as well as a post-training THEN/POST self-report.

Sample

Public assistance staff who participated in the training totaled 116. Videotaped interviews of 36 participants were randomly selected for analysis. Of these, 13 subjects had completed both the pre- and post-

videotapes (for a total of 26 interviews), while 10 had completed only the post. For the analysis of the interviewing effectiveness self-report, the primary analysis was of 35 subjects who filled out the PRE, the THEN, and the POST self-reports. A comparison group of 35 participants filled out the POST self-reports only. Seventy subjects were used in the analysis of knowledge test scores.

Results

Behavioral Ratings

The analysis of frequency of reflections, of frequency of attending behaviors, and of the global ratings on empathy, respect, and genuineness all indicated statistically significant changes during training. There appeared to be no stable change in frequency of questions. All four measures were investigated using a one-way analysis of variance. In each case, a comparison was made to check for the influence of pretesting on the posttest. This last point deserves additional discussion.

The pretest/posttest design, which is the basis for the following comparisons, is a weak design because it does not rule out the possibility that differences between pre- and posttest scores might be due to maturation, history, or the influence of the pretest on either the treatment or the posttest. Maturation is not a plausible explanation for pretest-to-posttest improvements because the participants were of widely differing ages and years of work experience. Similarly, historical events that might have occurred between the beginning and end of training could not explain pretest-posttest improvement because the training occurred during nine separate, 1-week-long sessions spread over the course of 1 year.

Because the videotape used for the pretraining assessment of ability was an integral part of training, there was no threat of a spurious effect on training or the posttraining assessment. However, in order to investigate whether the pretraining consent might have had an effect on training outcomes as measured by the posttraining videotape assessments, posttraining assessments were conducted on 10 subjects who were not requested to provide pretraining data. For all component measures, the posttest means were very similar for the two groups. For example, the mean posttraining global rating for the group that provided the pretraining assessment was 2.00, and for the group that was asked to provide only a posttraining assessment it was 2.01. Thus the use of the pretraining to posttraining videotape assessments in order to capitalize on the more powerful within-subjects design[13] appears justifiable.

Use of all types of reflections increased significantly after training. Reflections of content increased from an average of 1.2 occurrences prior to training to 2.9 after training, $F(1,13) = 6.7, p < .05$. Reflections

of feeling increased from an average of 0.4 occurrences prior to training to 1.5 after training, $F(1,13) = 12.1, p < .005$. Finally, simultaneous reflections of both content and feelings increased from no occurrences prior to training to an average of 0.8 after training, $F(1,13) = 11.1, p < .005$. Overall use of reflections increased from the average of 1.6 prior to training to 5.2 after training, $F(1,13) = 11.9, p < .005$).

There were also significantly more attending behaviors after training. Participants were observed briefly at eight intervals during the videotape to determine if they were exhibiting attending behaviors. Whereas an average of only .9 occurrences were observed prior to training, an average of 5.3 occurrences were recorded after training, $F(1,13) = 17.1, p < .001$.

There was also a significant difference, though not as dramatic, in the pre- and postglobal ratings of empathy, respect, and genuineness. Videotaped interviews prior to training were rated on the average at 1.95, or slightly hurtful, whereas interviews posttraining were rated on the average 2.0, neither helpful nor hurtful, $F(1,13) = 6.88, p < .02$). The global ratings are understated in that they are based on all statements and questions made by the worker. Since some worker communication is not meant to be helpful but only to elicit information, it is somewhat inappropriate to rate all communications on the global scale.

Another way of describing these data is to report the frequency of response. While there was, on the average, less than one helpful response every five interviews on prevideotapes, there were, on the average, more than 10 helpful responses every five interviews on the posttapes. There were no responses rated 3.0 or higher on the pretapes, while there was an average of more than one per interview on the posttapes. These differences in skill level are not as apparent in the average global ratings because of the large number of communications involved.

Knowledge test.—The knowledge test consisted of 35 items with an internal consistency reliability of .72 (Cronbach's alpha). Because it was possible that the knowledge pretest would interact with training to affect knowledge posttest scores, 70 participants were randomly assigned either to receive both a pretest and a posttest or to receive a posttest only. The conservative comparison is between the posttest of the 35 who received the posttest only and the pretest of the 35 who received both the pretest and the posttest. This between-subjects comparison indicates a significant training effect, $t(68) = 4.41, p < .001$. The within-subject comparison was also significant, $t(34) = 6.58, p < .001$. The means for the group receiving both pre- and posttests were 24 points and 30.2 points, respectively. The mean for the posttest-only group was 29.44 points.

Interviewing effectiveness self-reports.—This instrument provided information about how participants assessed their own interviewing ef-

fectiveness. In order to deal with some of the methodological problems involved in using pre-post self-report instruments, participants were randomly assigned to one of two data-collection schedules. One group filled out the assessment of interviewing effectiveness only after training and judged only their current posttraining ability (the POST score). The second group of participants responded both prior to training (the PRE score) and after training. The assessment after training for this group was so designed that participants retrospectively judged their ability before training (the THEN score) at the same time that they judged their current ability after training (the POST score). If there were no effects of filling out the PRE scales and the THEN scales on judgments of current ability, then the POST scores for the two groups would be about the same. Mean POST scores were not significantly different, $F(1.65) \sim 1$, N.S. With 15 representing a perfect score and 105 representing the lowest possible judgment of one's ability, the average POST score for group 1 was 36.7 and for group 2 was 35.9. Given that the rating of PRE and THEN ability apparently did not distort group 2 participants' ratings of POST ability, the more interesting comparison was whether group 2 participants viewed themselves as more able before or after training. Analysis of PRE, THEN, and POST scores for group 2 participants yielded significant differences between PRE and THEN scores and between THEN and POST scores, but not between PRE and POST scores. The mean on the PRE measure was 38.8, on the POST measure was 35.9, and on the THEN measure was 45.2, $F(2,32) = 11.4, p < .001$ (higher scores represent lower perceived ability). These results suggest that as learning took place, participants got a better understanding of what they did not know and judged their pretraining ability lower after training than before training. Additionally, these results suggest that after training, participants perceived their skill to be improved considerably. The PRE to POST comparison was invalid, apparently because a response shift had taken place. It was the THEN to POST comparison that indicated how participants judged the effects of training.[14]

Participant evaluations.—Participants responded quite favorably to the workshop. Using a seven-point Likert-type agree/disagree scale (1 = strongly disagree, 7 = strongly agree), average ratings on four items were as follows: "content was interesting," 6.3; "objectives were made clear," 6.3; "trainer was prepared," 6.5; and "training was helpful at work," 6.2.

Participant evaluations indicated that the skills learned were very relevant to their jobs, that they perceived the training to be of high quality, and that instructional methods were appropriate, despite some discomfort with the videotaping. Overwhelmingly, participants requested that a follow-up training program be developed. Criticism of the program centered on the long hours and intensity of the training experience.

Discussion

The results of the CIE I evaluation provided more complete information than had been available on any program previously conducted by the agency. The information suggested positive gains for participants and was valuable to the OCSWE staff and the Division of Family and Children Services.

Participants showed gains in knowledge and perceived their level of skill use to have improved. The results of the behavioral ratings suggest caution because the magnitude of the skill gains, albeit statistically significant, was of unknown practical significance. For instance, how much more effective is an interview in which a worker uses 2.9 reflections of content than one in which only 1.2 are used? Are 1.5 reflections of feeling indicative of a significantly higher ability to relate to client feelings than .4 such occurrences? Are participants really able to be more effective if "helpful" responses occur an average of twice an interview rather than less than once every five interviews? Furthermore, limitations in the design left unanswered questions regarding retention of even these small gains. Would they deteriorate quickly? If so, the training represented a considerable investment for a small amount of possibly fleeting improvement. These questions remain unanswered. However, there were valuable conclusions resulting from this effort.

First, the evaluation was successfully integrated into the training design without disruption. The measurement strategies, for the most part, served two purposes. The videotaped interviews and knowledge tests provided evaluation data as well as opportunities for practice and feedback within the training itself. The feedback component was very popular with participants, who commented that feedback on progress and learning is missing from many training programs. Although scheduling these tapings required additional coordination and forced some down time at the beginning of the program, the trainees did not seem to mind starting with the videotaping, nor was there any evidence that this hindered the formation of group cohesion or trainer-trainee relationships.

Second, Mitchell's observations on the time required to learn empathy had relevance to CIE.[15] The training program Mitchell studied involved six skills (core conditions). The CIE I program attempted to train in seven skills and three core conditions. Although the total training time was 30 hours, the time per skill was only an average of 3 hours. The small gains observed might be attributed to poor instructional design, trainer ineffectiveness, or might be a function of the small amount of actual time devoted to each skill. The evaluation clearly pointed to weaknesses in some skill modules, and these were subsequently revised.

The authors believe that substantial attention must be paid to learning as a function of time. Becoming proficient in the use of interpersonal

helping skills involves learning new behavioral cues, incorporating new responses, and achieving genuineness in the process. It also involves unlearning dysfunctional cues and responses. Such learning and unlearning require considerable practice under varied conditions with appropriate feedback. Had CIE I focused on mastery of a smaller number of skills, the gains may have been more significant.

While positive gains were found in the CIE program, the results of the evaluation also raise questions about expectations for short-term, in-service training programs. One must question whether a few days of workshop training can prepare staff to be interpersonally effective and provide an adequate basis for practice. Only by more thoroughly evaluating short-term training programs in terms of actual changes in participant knowledge, skill, and attitudes can policymakers and instructional designers have information necessary to make sound decisions on such issues as training content, instructional activities, length of time, and other factors. Further studies are needed to document retention or loss of skills and knowledge over time. Particular emphasis should be placed on the role of unlearning previous behavioral cues and patterns when training focuses on well-imbedded behaviors such as interpersonal skills.

Notes

1. E. K. Marshall, J. W. Charping, and W. J. Bell, "Interpersonal Skills Training: A Review of the Research," *Social Work Research and Abstracts* 15 (1979): 10–16; S. J. Danish, A. R. D'Augelli, and G. W. Brock, "An Evaluation of Helping Skills Training: Effects on Helpers' Verbal Responses," *Journal of Counseling Psychology* 23 (1976): 259–66.

2. Thomas D. Morton, "Evaluation Methodology in Continuing Education: A Review of Reported Evaluations between 1970 and 1980" (paper presented at the Conference on Training in the Human Services, Albany, N.Y., October 1981).

3. S. Mitchell, *A Process Study of Systematic Interpersonal Skills Training* (Ph.D. diss., University of Georgia, Athens, 1974).

4. T. D. Morton and P. D. Kurtz, "Design Considerations for Short-Term Continuing Education Programs," *Journal of Education for Social Work* 20 (Winter 1984): 42–49.

5. A. E. Ivey and J. Authier, *Microcounseling* (Springfield, Ill.: Charles C. Thomas, 1978).

6. C. R. Rogers, "The Necessary and Sufficient Conditions of Therapeutic Personality Change," *Journal of Consulting Psychology* 21 (1957): 95–103; S. P. Shinke, T. E. Smith, L. D. Gilchrist, and S. E. Wang, "Measuring the Impact of Continuing Education," *Journal of Education for Social Work* 17 (Winter 1981): 59–64.

7. Ivey and Authier.

8. T. D. Morton and E. W. Lindsey, "Toward a Model for Interpersonal Helping Skills Use and Training in Public Welfare Practice," *Journal of Continuing Social Work Education* 4 (Winter 1986): 18–24.

9. George Gazda, Frank Asbury, F. Balzer, W. Cilders, and R. Walters, *Human Relations Development* (Boston: Allyn & Bacon, 1973).

10. A. Benjamin, *The Helping Interview* (Boston: Houghton, Mifflin & Co., 1974); G. Egan, *The Skilled Helper* (Monterey, Calif.: Brooks/Cole Publishing Co., 1975); Ivey and Authier.

11. D. T. Campbell and J. C. Stanley, "Experimental and Quasi-Experimental Designs for Research on Teaching," in *Handbook of Research on Teaching*, ed. N. L. Gage (Chicago:

Rand McNally & Co., 1963); T. D. Cook and D. T. Campbell, *Quasi-Experimentation: Design and Analysis Issues for Field Settings* (Chicago: Rand McNally & Co., 1979).

12. G. Howard, "Response-Shift Bias: A Problem in Evaluating Interventions with Pre/Post Self-Reports," *Evaluation Review* 4 (1980): 93–106.

13. R. Kirk, *Experimental Design: Procedures for the Behavioral Sciences* (Belmont, Calif.: Brooks/Cole Publishing Co., 1968).

14. Howard.

15. Mitchell (n. 3 above).