AN ANALYSIS OF SOCIAL SKILLS GENERALIZATION IN TWO NATURAL SETTINGS

R. M. FOXX, MARTIN J. McMORROW, RON G. BITTLE, AND
JAYNE NESS

ANNA MENTAL HEALTH AND DEVELOPMENTAL CENTER AND
SOUTHERN ILLINOIS UNIVERSITY

The interactional behavior of two groups of elderly mentally retarded residents of a community facility was measured in two generalization situations before, during, and after one group received social skills training. The training group received social skills training within a game format, whereas the contrast group simply played a game with no emphasis on interactional behavior. Results suggested that generalization to natural interactional situations may be delayed following training and that it is more likely in some situations (i.e., with trained peers) than others (i.e., in the presence of untrained peers).

DESCRIPTORS: social skills, educational game, generalization, mentally retarded, elderly, social validity

The generalization of appropriate social behavior from training settings to those in which subjects normally interact has been a controversial and complex issue in social skills research. Whether generalization occurs, with whom, what assessments will demonstrate the phenomenon, and when to anticipate it are just a few of the questions that have been raised. The lack of consistent findings regarding these questions has led to conceptual reformulations of social skills (e.g., McFall, 1982), skepticism concerning the social validity of many training efforts (Berler, Gross, & Drabman, 1982), and even a call for a moratorium on training research until useful assessments are made (Matson, Esveltd-Dawson, & Kazdin, 1983).

In an attempt to address these questions, our social skills research has used two types of generalization assessments: (a) analogue or confederate-mediated tests (Foxx & McMorrow, 1985; Foxx, McMorrow, Bittle, & Fenlon, 1985; Foxx, McMorrow, & Mennemeier, 1984a; Foxx, McMorrow, & Schloss, 1983; Foxx, McMorrow, Storey, & Rogers, 1984b), and (b) nonmediated observations conducted under more natural conditions

(Foxx et al., 1984a). Although this research has yielded encouraging results, it has failed to provide definitive answers to the above questions for two reasons. First, conclusions regarding generalization that are based on analogue tests can never be definitive because such tests cannot be considered a direct measure of the phenomenon (i.e., a training-related change in social behavior in settings where subjects normally interact). Second, the design used in our natural assessment (Foxx et al., 1984a) failed to demonstrate a functional relation between social skills training and the subjects' improvements in the generalization setting because the improvements were not evident until weeks after training had ended.

Nevertheless, Foxx et al. (1984a) raised the possibility that generalization following the initiation of social skills training may be delayed since the design did not rule out the possibility that the improvements during follow-up were related to training. Furthermore, individual data revealed a number of intersubject interactional consistencies. For example, when the behavior of one member of a group was primarily inappropriate, the others' behavior also tended to be inappropriate. These patterns suggested that an individual's social behavior was highly dependent on the behavior of others. These issues raise the possibility that previous social skills research has failed to demonstrate

We thank Sarah Fenlon and Maya Roeder for their assistance.

Reprint requests should be sent to R. M. Foxx, Anna Mental Health and Developmental Center, 1000 N. Main St., Anna, Illinois 62906.

generalization because the assessments were not continued long enough, or because the subjects were observed in an environment characterized by consistent and/or long-term patterns of inappropriate interaction.

One goal of our study was to test the hypothesis that the generalization of trained social skills to a natural environment may be delayed for persons with long histories of inappropriate social behavior. This was done by (a) having long-term residents of a group home for mentally retarded persons serve as subjects, (b) assessing the social behavior of two groups in the generalization settings, but withholding training for one group, and (c) conducting natural generalization assessments before, during, and after the social skills training. A second goal was to determine how the presence of untrained peers in a generalization setting influenced the trained subjects' interactional behaviors. This was done by observing the group's interactional behavior in two natural settings; one that involved only the group and another that included untrained peers.

METHOD

Subjects

Six elderly, mentally retarded, female residents of a community-based facility for physically and mentally disabled adults served as subjects. They were selected on the basis of staff referrals, time in residence (more than 10 years at the facility), and low scores on pretraining interviews (less than 50% correct). The subjects were matched into one of two groups ($n = 3$ per group) on the basis of their responses to 24 of the training situations in an individual interview conducted by the fourth author prior to the start of baseline observations. During the interviews, subjects were presented with half of the situations from the General Social Skills curriculum of the *Stacking the Deck* program (Foxx & McMorro, 1983) but no feedback was provided. Their responses were evaluated using the program's scoring criteria (see Foxx *et al.*, 1984a). Table 1 presents individual demographic information and interview scores.

Table 1
Individual and Group Demographic Information and
Pretraining Social Skills Interview Scores^a

	Age	I.Q. ^b	Years in residence	Interview score ^c
Training Group				
S1	69	55	17	41.7
S2	63	26	15	8.3
S3	76	60	18	25.0
<i>M</i>	69	47	16.7	25.0
Contrast Group				
S1	68	50	15	29.2
S2	58	90 ^d	18	39.1
S3	63	70	12	8.3
<i>M</i>	63	70	15.0	25.5

^a All subjects could speak in complete sentences.

^b Scores based on either the Vineland or Wechsler test.

^c Percent correct on 24-item pretraining interview and basis for matching subjects into groups.

^d Subject had a dual diagnosis.

Experimental Design

Both groups' interactional behavior was assessed in two natural settings; one with the group only and one with the group plus untrained peers. After baseline observations, the training group received the General Social Skills program outlined in Foxx *et al.* (1983), while the contrast group simply played a standard "Sorry" game (see procedures below). After the game condition, follow-up assessments were conducted for both groups in both settings.

Settings

All games were played in a lounge area of one of the facility's dormitory suites. *Group only* assessments were conducted in a different lounge area. *Group plus untrained peers* assessments were conducted in the facility's dining room during lunch.

Data collection

Target behaviors. The six social interaction components (i.e., compliments, social interaction, politeness, criticism, social confrontation, and questions/answers) described in Foxx *et al.* (1983) were targeted. Each component area was broken down into "actor" (e.g., giving a compliment) and "reactor" (e.g., responding to a compliment) cat-

egories. During training, each subject's responses to 48 training situations (8 situations per component, 4 per category) were scored as correct or incorrect using the criteria specified in Foxx et al. (1983).

Responses during the generalization assessments also were scored as in Foxx et al. (1984a), except that, instead of time sampling naturally occurring interactions, all intelligible verbalizations that occurred during the first 10 minutes of each assessment were scored. This was accomplished by audiotaping each session and later transcribing each verbalization. Each response was evaluated as either appropriate or inappropriate by assigning it to a component area (e.g., compliments) and category (i.e., actor or reactor), and then using the scoring criteria for that area. Additional scoring rules included: (a) scoring any subject's failure to respond to an action as inappropriate unless another group member or peer responded; (b) excluding the staff/subject interactions that occasionally occurred at lunch; (c) including only personalized statements in the compliments, criticism, and social confrontation component areas (e.g., "Betty, you look nice today."); and (d) scoring any verbal interruption of another's ongoing verbalization as inappropriate.

Procedures

Group only generalization assessments were conducted on an average of once every 5 days when no game was scheduled. Each group was simply escorted to the lounge and asked to "sit and talk" for 10 minutes. Scoring was discontinued if a group member left the lounge before 10 minutes had elapsed. No attempt was made to coerce an individual to stay since we wanted the situation to be as natural as possible. No staff or training personnel were present.

Group plus untrained peers assessments were conducted almost daily at lunch. Between 9 and 13 residents sat at two long tables at each of two meal shifts. Each group was assigned to a different shift and sat at one table to facilitate audiotaping. One facility staff member was always present and started the audiotaping when everyone was seated

and ended it when the first subject completed the meal (only the first 10 min of the tape were analyzed). The social skills trainer was never present. All facility staff were kept naive regarding the specific aspects of the study.

Observations of both groups were conducted in both generalization settings before, during, and after the social skills training or standard game conditions. Baseline assessments began 2 weeks prior to the game conditions. Following these assessments, the training group received the General Social Skills program described in Foxx et al. (1983). This program incorporates a special set of cards that depict the 48 social situations into the "Sorry" table game (Parker Brothers, Salem, Massachusetts). The cards were prearranged so that each subject received a different set of 12 situations each game. Hence, four games were required before a subject was exposed to all 48 situations. During baseline games, subjects move their playing pieces independently of their responses to the game cards. Snacks were provided after each game. During training, subjects moved their pieces only when their responses were appropriate. The trainer (fourth author) provided specific positive or corrective feedback after each response and modeled appropriate responses on her turns. The subjects continued to receive a snack after each game, and those who met individualized, prespecified, and changing criteria earned preselected rewards. The trainer simply played the usual version of "Sorry" with the contrast group with no emphasis on their interactional behavior. A snack was provided following each game.

Both groups' game sessions were conducted during a 5-week period on the same days and by the same individual (i.e., both participated in 20 games). After the game conditions, follow-up generalization assessments were conducted once a week for 15 weeks in the group only condition and four times weekly for 6 weeks in the group plus untrained peers condition (lunchtime). During follow-up, the trainer's only involvement at the facility was to escort subjects to the group only assessments and pick up the recordings from lunch.

Reliability. Randomly selected transcripts from

the training games and generalization assessments were independently scored by the trainer and a separate rater and percent interobserver agreement was calculated. Mean reliability for the training group was 95% during 6 (27%) of the group only assessments, 99% during 13 (25%) of the group plus untrained peers assessments, and 91% during 6 (30%) of the social skills training games. Mean reliability for the contrast group was 93% during 6 (27%) of the group only assessments and 94% during 13 (25%) of the group plus untrained peers assessments.

RESULTS

Training Results

All three subjects' mean percentage of correct responses to the game situations increased during training, with the smallest individual gain being 23% and the largest 48%. The group averaged 16% during baseline games and reached 75% correct after 10 training games.

Group Only Generalization Assessments

Of the group only assessments, 66% were 10 minutes in length and 89% lasted between 7 and 10 minutes. Only five sessions lasted less than 7 minutes (see Figure 1). There was no correlation between session length and either percentage or rate of appropriate responding.

Figure 1 shows that during baseline the training group averaged 42.5 interactions per session, of which 26.5 (62.4%) were appropriate. The contrast group averaged 50.0 interactions, of which 34.5 (69.0%) were appropriate. During baseline, training group S3 never engaged in an appropriate interaction and contrast group S3 never engaged in an inappropriate interaction. The interactional patterns did not change much during the game conditions. Percentages of appropriate interactions were remarkably similar for both groups (69.9% and 71.0% for the training and contrast groups, respectively) and did not change appreciably from baseline.

However, clear within- and between-group differences appeared during follow-up. The most im-

Figure 1. Each subject's appropriate and inappropriate interactions during the group only generalization assessments. Training group data are in the upper two panels and contrast group data are in the lower two. Sessions during weeks 2, 7, and 14 for the training group and 13 and 21 for the contrast group lasted less than 7 minutes.

portant was that training group members S1 and S2 showed a substantial increase in their number and percentage of appropriate interactions and a great deal of intersubject consistency (i.e., a characteristic pattern of interaction with each other) that seemed to increase as follow-up progressed. Analysis of the audiotapes revealed that (a) the majority of their interactions occurred within the

Figure 2. Each subject's appropriate and inappropriate interactions during the group plus untrained peers generalization assessments. Training group data are in the upper two panels and contrast group data are in the lower two. Each weekly data point represents the mean number of interactions during four lunchtime meals that week.

social interaction and questions/answers component areas; (b) both initiated approximately half of the interactions; and (c) failures to respond to each other became less frequent as the follow-up progressed. The training group's mean percentage of appropriate interactions during follow-up increased to 81%, whereas the contrast group's mean of 67.1% closely approximated their baseline mean.

The training group's interactional improvements during follow-up seem particularly significant given that their mean number of interactions per session, 86.1%, more than doubled either of the two previous conditions, their frequency of inappropriate interactions did not change appreciably, and the contrast group showed no clear changes in either appropriate or inappropriate interactions. Interestingly, the only intersubject consistency in the contrast group was between S1 and S2 on their inappropriate interactions (see Figure 1).

Group plus Untrained Peers Generalization Assessments

Of the 104 lunchtime assessments, 84% were 10 minutes in length, 94% lasted between 7 and 10 minutes, and only seven sessions lasted less than 7 minutes. Figure 2 shows that during baseline, the training group averaged 11.2 interactions per observation of which 2.6 (23.2%) were appropriate, whereas the contrast group averaged 4.4 and 2.1 (47.7%) interactions, respectively. The only effects of game participation were shown by training group S1 whose appropriate interactions increased markedly during weeks 6 and 7 (see Figure 2). During this period, training group S2 and S3's inappropriate interactions increased because they failed to respond to S1's appropriate interactions. Although very few responses followed S1's appropriate initiations, she persisted in her attempts to interact until week 13. There were no apparent effects in the contrast group.

Training Group S3

Training group S3 participated in only two appropriate interactions during the entire assessment and all of her inappropriate interactions were failures to respond. This lack of responsiveness was unexpected because her behavior during the pre-training interview and in training resembled that of her peers. That is, she scored 25% correct in the interview and averaged 75% correct during the training sessions. In an attempt to isolate the cause of these differences, the group only condition was re-created after the study with two normal women taking the place of S3's peers. The two normal

women conversed with each other during five 10-min sessions and initiated an interaction with S3 once every 2 min. This rate of interaction approximated S3's mean frequency of failures to respond during the follow-up condition in the group only assessment. Although S3 still did not initiate an interaction, she responded appropriately to 53% of the verbalizations directed to her.

DISCUSSION

This study yielded two findings regarding the generalization of a set of norm-based social skills (i.e., the targeted behaviors had been developed by polling normal individuals; see Foxx *et al.*, 1983) to some of the situations in which mentally retarded individuals normally interact. First, the group only assessment demonstrated that the use of such skills may be delayed, since two of the three trained subjects used them during the 15-week follow-up but failed to do so in the generalization assessments conducted during training. Second, the training and generalization results suggested that the subjects were more likely to use these skills in some situations than in others. In the training group (a) all three subjects used the trained behaviors in the game situation when they interacted with the trainer, (b) two of three used the trained skills when they interacted with each other, (c) one subject (S3) interacted appropriately with two normal individuals, but failed to interact with trained or untrained peers, and (d) the use of trained skills in the presence of untrained peers was either virtually nonexistent (S2 and S3) or sporadic (S1).

These findings may help explain why previous studies have yielded conflicting conclusions regarding social skills generalization. For example, if generalization of social interaction does not correspond to the onset or implementation of training, we might not expect it in studies that conduct assessments during or immediately following training. In addition, if generalization of norm-based skills is more likely in some situations (e.g., with trained peers or normal individuals) than others (e.g., with untrained peers), we would not expect it in studies

that conduct assessments in situations characterized by other types of behavior from untrained individuals (e.g., those in institutional dayrooms). Furthermore, the rapidity of the effects may be related to the kind of situation in which generalization is assessed. These explanations suggest that future research should pay special attention to when, how long, where, and with whom generalization assessments are conducted.

These findings also have implications for subsequent conceptualizations of socially skillful behavior. Specifically, because the subjects' use of trained behaviors increased when they interacted with others who exhibited similar behaviors (i.e., trained peers or normal individuals), we would assume that these behaviors were reinforced. On the other hand, it appears that the trained behaviors were not reinforced by peers who did not exhibit them. In fact, given the patterns of inappropriate interactional behavior in both groups and between some subjects (e.g., contrast group S1 and S2, see Figure 1), we would assume that other types of interactional behaviors were reinforced. Thus, it appears that different behaviors were effective (i.e., reinforced) in the different situations. This conceptualization suggests that socially skillful behavior may not always be appropriate. Thus, future research should clearly distinguish between skillful (or effective) interactional behavior and appropriate (or norm-based) interactional behavior.

This distinction may be especially useful in determining the social validity of a program's targeted skills. Put simply, the extent to which trained skills generalize to a particular setting might be viewed as a direct measure of their social validity in that setting. If we follow this logic, the skills taught in this study were socially valid in the facility's lounge but not in its lunchroom.

In conclusion, a careful consideration of situational differences in social behavior may aid subsequent training efforts and promote generalization. An assessment of these differences could help determine when and what type of training is needed. For example, an individual may not necessarily have social skills deficiencies simply because he or she does not use a particular set of behaviors in an

environment where they are not effective. Such skill deficiencies can only be identified by assessing whether the individual uses the trained skills in an environment where they are effective. If this assessment revealed deficiencies, the decision of whether and how to remedy them would require an examination of the individual's current and anticipated interactional environment. For example, if the individual lives in an environment where a particular set of behaviors would not be effective, our results suggest that training such behaviors would be futile unless many individuals in that environment were trained. Conversely, if the environment supports a particular type of behavior but the individual does not display it, more effective behaviors should be trained as quickly as possible. Finally, if the individual's placement goal is an environment where a particular set of behaviors is effective but he or she lacks it, then training should be given shortly before placement.

REFERENCES

- Berler, E. S., Gross, A. M., & Drabman, A. S. (1982). Social skills training with children: Proceed with caution. *Journal of Applied Behavior Analysis*, *15*, 41-53.
- Foxx, R. M., & McMorrow, M. J. (1983). *Stacking the deck. A social skills game for retarded adults*. Champaign, IL: Research Press.
- Foxx, R. M., & McMorrow, M. J. (1985). Teaching social skills to mentally retarded adults: Follow-up results from three studies. *The Behavior Therapist*, *8*, 77-78.
- Foxx, R. M., McMorrow, M. J., Bittle, R. G., & Fenlon, S. J. (1985). Teaching social skills to psychiatric inpatients. *Behaviour Research and Therapy*, *23*, 531-537.
- Foxx, R. M., McMorrow, M. J., & Mennemeier, M. (1984a). Teaching social/vocational skills to retarded adults with a modified table game: An analysis of generalization. *Journal of Applied Behavior Analysis*, *17*, 343-352.
- Foxx, R. M., McMorrow, M. J., & Schloss, C. N. (1983). Stacking the deck: Teaching social skills to retarded adults with a modified table game. *Journal of Applied Behavior Analysis*, *16*, 157-170.
- Foxx, R. M., McMorrow, M. J., Storey, K., & Rogers, B. M. (1984b). Teaching social/sexual skills to mentally retarded adults. *American Journal of Mental Deficiency*, *89*, 9-15.
- Matson, J. L., Esveldt-Dawson, K., & Kazdin, A. E. (1983). Validation of methods for assessing social skills in children. *Journal of Clinical Child Psychology*, *12*, 174-180.
- McFall, R. M. (1982). A review and reformulation of the concept of social skills. *Behavioral Assessment*, *4*, 1-33.

Received September 5, 1985

Final acceptance March 18, 1986