JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1976, 9, 127-139 NUMBER 2 (SUMMER) 1976

# THE SOCIAL VALIDATION AND TRAINING OF CONVERSATIONAL SKILLS[1]

NEIL MINKIN, CURTIS J. BRAUKMANN, BONNIE L. MINKIN, GARY D. TIMBERS,[2] BARBARA J. TIMBERS,[2] DEAN L. FIXSEN, ELERY L. PHILLIPS, AND MONTROSE M. WOLF

UNIVERSITY OF KANSAS

Three reliably measured components of conversation—questioning, providing positive feedback, and proportion of time spent talking—were identified and validated as to their social importance. The social validity of the three conversational behaviors was established with five female university students and five female junior-high students. Each was videotaped in conversations with previously unknown adults. The conversational ability of each girl was evaluated by a group of 13 adult judges who viewed each tape and rated each conversant "poor" to "excellent" on a seven-point rating scale. The average ratings of the girls correlated at $r = 0.85$ with the specified behavioral measures. These procedures were replicated with additional subjects and judges and yielded a correlation of $r = 0.84$. The high correlations between ratings and the objective measures suggested that the specified conversational behaviors were socially important aspects of conversational ability. Employing a multiple-baseline design across the behaviors of asking questions and providing positive feedback, an attempt was made to train four girls who used these behaviors minimally to engage in the behaviors in conversations with adults. Adult judges were again employed to rate randomly selected samples of the girls' skills in pre- and posttraining conversations. The average ratings of the girls before training were lower than both the university girls and the junior high-school girls. After training, the girls' conversational abilities were rated substantially higher than those of their junior high-school peers. These rating data validated the benefits of the training and the social importance of the behavioral components of questions and feedback in conversation. The authors suggest that it may be necessary for traditional behavior analysis measurement systems to be supplemented by social-validation procedures in order to establish the relationship between "objectively" measured behaviors and complex classes of behavior of interest to society.

DESCRIPTORS: social validation procedures, conversational behavior, social interaction behaviors, predelinquents

The effectiveness of applied behavior analysis depends on careful specification and measurement of the behavior of interest. Many behaviors dealt with by behavior analysts are easily specified and measured *e.g.,* working arithmetic problems correctly (Felixbrod and O'Leary, 1973) or climbing on a climbing apparatus (Harris, Wolf, and Baer, 1964). However, some behaviors are more complex and difficult, especially socially important behaviors that include numerous component parts. These behaviors are often described in vague generalities, which do not

[1]Portions of these data were presented at the eighty-first Annual Convention of the American Psychological Association, Montreal, Canada, September 1973. This manuscript is based upon a thesis submitted by the senior author to the Department of Human Development, University of Kansas, in partial fulfillment of the requirements for the M.A. degree. This research was supported by grants MH-13664 and MH20030 from the National Institute of Mental Health (Center for Studies in Crime and Delinquency) to the Bureau of Child Research and the Department of Human Development of the University of Kansas. The authors express their appreciation to Hector Ayala, Patrizia Braukmann, Richard Goldstein, Steven King, Kathryn Kirigin, John Lutzker, Dennis Maloney, Howard Rosenfeld, Larry Russell, Mitchell Taubman, John Werner, and a number of others who contributed to this study. Reprints may be obtained from Neil Minkin, Achievement Place Research Project, Department of Human Development, University of Kansas, Lawrence, Kansas 66045.

[2]Now at the Western Carolina Center, Morganton, North Carolina.

provide a basis for measurement. For example, empathy is considered to be an important characteristic of an effective counsellor (Truax and Carkhuff, 1967), yet what exactly constitutes empathy and thus, what behaviors should be taught to counsellor trainees, is debatable (Coleman, 1964; Haase and Tepper, 1972; Smith, 1973). To quantify empathy, researchers have attempted to specify the behavioral components and to validate their importance through the ratings of relevant judges *i.e.,* experienced counsellors.

Haase and Tepper (1972) asked experienced counsellors to rate the empathy level of a counsellor who was modelling various specified behavioral components of empathy on videotaped segments of simulated counselling situations. The results indicated that several nonverbal behaviors, including eye contact, bodily orientation toward the client, and the distance of the counsellor from the client correlated highly with the ratings. This outcome indicated that these components were "valid" aspects of empathy according to the judgement of experienced counsellors. The specification of behavioral components and validation of their importance by relevant judges is a procedure that might be used to define other complex social interaction skills.

Finding a high positive correlation between the specified behaviors and the ratings of skill levels by relevant judges does not rule out the possibility that some important behavioral components of the skill remain unspecified. In addition, frequency levels, duration levels, and critical interaction patterns of the behaviors may be important variables. Given these possibilities, training that increases levels of the specified and validated behaviors may not increase ratings of the skill level by relevant judges. Validation that training did increase judged skill level might be obtained by asking relevant judges to rate an individual's skill level both before and after training.

Thus, the specification and training of complex social behaviors that involve subjective dimensions seems to require four steps: (1) specification of the potentially relevant behavioral components, (2) social validation of the importance of each of the behavioral components, (3) training of the components, and (4) social validation that increases in the specified behavioral components resulted in increased level of judged skill. The purpose of the present research was to carry out these steps in a training program designed to improve the conversational skills of predelinquent girls.

## I. SPECIFICATION AND SOCIAL VALIDATION OF CONVERSATIONAL BEHAVIORS

Two sets of conversational samples were obtained to permit the reliable specification and social validation of some important behavioral components of conversation. The first set was used to identify and to measure conversational behaviors, to provide normative information on the conversational behavior of junior-high and university females, and to determine the relationship between the conversants' behavior and ratings of conversational skill by relevant judges.

### Sample I

#### *Subjects*

Five junior high-school students and five university students, all female, ranged in age from 18 to 20 yr. The junior-high students were in the eighth and ninth grades and the university students were freshmen and sophomores. The junior-high girls volunteered in response to an announcement made in a study hall by the vice-principal of a local junior high school. The university girls responded to an announcement made in a sorority house by one of the sorority sisters.

#### *Setting and Apparatus*

The setting was a 3.6 by 4.2 m room that contained two chairs positioned at an angle of 45 degrees with respect to one another. A video

camera and microphone, placed approximately 1.5 m in front of the two chairs, were connected to a Sony 2200 videotape recorder and monitor located in an adjoining room not visible from the chairs.

*Procedure*

Each of the 10 subjects was videotaped in two 4-min conversations, producing a total of 20 sample conversations. In each conversation, the subject and an adult previously unknown to the subject were alone in the room. The sequencing of conversations was arranged so that no subject would have two consecutive conversations. Each subject conversed with a different male or female adult during each conversation.

On entering the room, the conversants were asked to be seated and given the following verbal instructions by one of the experimenters:

> "We would like you to speak with each other for a short period of time. You may talk about anything you wish. You will be told when to begin and when to stop."

The videotape unit was then turned on. The experimenter said, "You may begin now" and left the room. At the end of 4 min, the videotape unit was turned off, the experimenter re-entered the room, informed the conversants that time was up, and thanked them for participating. Each conversant then received $2.00 for having participated.

*Definitional specification, recording, and reliability.* After informally reviewing the 20 videotaped sample conversations, the experimenters noted that the university students had asked more questions and had given more positive feedback than the junior-high students. In some conversations, one person spoke a great deal or very little. Based on these informal observations, the conversational behaviors of questioning, positive feedback, and time talked were reliably specified using the following procedure.

Two or more naive observers simultaneously viewed the videotaped conversations. They were given a written definition of the behavior and written instructions on how to record the behavior. An interval procedure was used to record conversational questions and positive conversational feedback by dividing each of the 4-min conversations into 24, 10-sec intervals. An audiotape, which signalled the beginning and end of each interval, was synchronized with the videotape for each conversation. The observers were instructed to score an occurrence in each 10-sec interval in which the behavior being observed occurred at least once. The written definitions of a conversational question and positive conversational feedback were modified several times until naive observers could agree at least 85% of the time that the behavior did or did not occur. To control for observer bias (Arrington, 1943), new observers were employed each time the written behavioral definition was changed. The time-talked measure was recorded by counting the cumulative number of seconds spoken by each of the conversants. This definition proved to be immediately reliable.

A "conversational question" was defined to include: (a) any command by the subject, (b) any question by the subject, (c) any "question of clarification" by the subject, and (d) any statement by the subject that in effect functions as a question.

Examples: following a statement by the other conversant, such as "I go to K.U.", an example of (a) (above) would be "Tell me more about that.", (b) (above) "How long have you gone there?", (c) (above) "K.U.?", or "Oh, really?", or "You do?", and (d) (above) "So you are in college." or "In other words you are a student.".

"Positive conversational feedback" was defined as a brief utterance of no more than three words that indicated that the subject either (a) approves, (b) concurs, or (c) understands what the other conversant is saying or has just said.

For example, if the other conversant were to say: "I think blue is the best color", examples of (a) (above) would be "That's nice", "good", or "interesting", (b) (above) "I agree", "I know", "mm-hmm", and "right", (c) (above) "oh", or

"hmm", or "blue" (a repetition of part of what was said unless intoned as a question). This utterance can be neither a "conversational question" nor a response to a "conversational question", and may or may not be directly followed by further utterances by the subject.

The reliability of the measurement system was assessed 128 times throughout the study. Four types of reliability computations were used: point by point, occurrence, nonoccurrence, and gross. Point-by-point agreement (Wolf and Sherman, Note 1) was calculated by dividing the total number of intervals of agreement by the total number of intervals of agreements and disagreements. Agreement as to occurrence (Bijou, Peterson, and Ault, 1968) was calculated by dividing the total number of intervals in which there was agreement that the behavior occurred by the total number of intervals in which there was agreement and disagreement that the behavior occurred. Agreement as to nonoccurrence was calculated by dividing the total number of intervals in which there was agreement that the behavior did not occur by the total number of intervals in which there was agreement and disagreement of the nonoccurrence of the behavior. The gross method (Wolf and Sherman, Note 1) of determining reliability was also used to compare the totals of the observers' observations. The gross per cent agreement was calculated only for the time-talked behavior by dividing the larger total into the smaller total.

Agreement on the recording of conversational questions was assessed on 48 occasions. Mean point-by-point agreement was 94% (range: 83% to 100%). Mean occurrence agreement was 90% (range: 75% to 100%), and mean nonoccurrence agreement was 91% (range: 66% to 100%).

Reliability on the recording of positive conversational feedback was assessed 50 times. Mean point-by-point agreement was 92% (range: 80% to 100%). Mean occurrence reliability was 89% (range: 66% to 100%), and mean nonoccurrence agreement was 90% (range: 66% to 100%). Interobserver reliability for time talked was assessed 30 times. The mean gross agreement was 98% (range: 88% to 100%).

## RESULTS

During the 24 intervals of a 4-min conversation, the university subjects averaged 7.7 intervals in which they asked at least one conversational question, 7.4 intervals containing at least one instance of positive conversational feedback, and spoke an average of 128 sec. The junior-high subjects averaged 0.8 intervals containing a conversational question, 3.2 intervals of positive conversational feedback, and spoke an average of 100.4 sec.

### *Social Validation*

*Judges and setting.* Adult residents of the subjects' local community observed the 20 sample conversations and rated the conversational ability of each subject. The purpose was to determine whether the specified behavioral components were viewed as socially important variables of conversation.

The 13 adults who volunteered to serve as judges were seven males and six females. They ranged in age from 19 to 51 yr, mean age, 26. Nine judges were full or part-time university students at the graduate or undergraduate level. The four other judges included a gas-station attendant, two homemakers, and a Pinkerton guard. The judges observed and rated the conversational ability of the subjects in a university conference room where two Sony series 2200 videotape monitors were used to display the taped conversations.

*Procedures.* The 20 sample conversations were arranged in random order, with the provision that no subject would appear in two consecutive conversation sequences. Finally, all conversations were transcribed onto two videotapes, each containing 10 conversations in the randomly determined order.

Two groups of judges consisting of six or seven people viewed the tapes. The first group

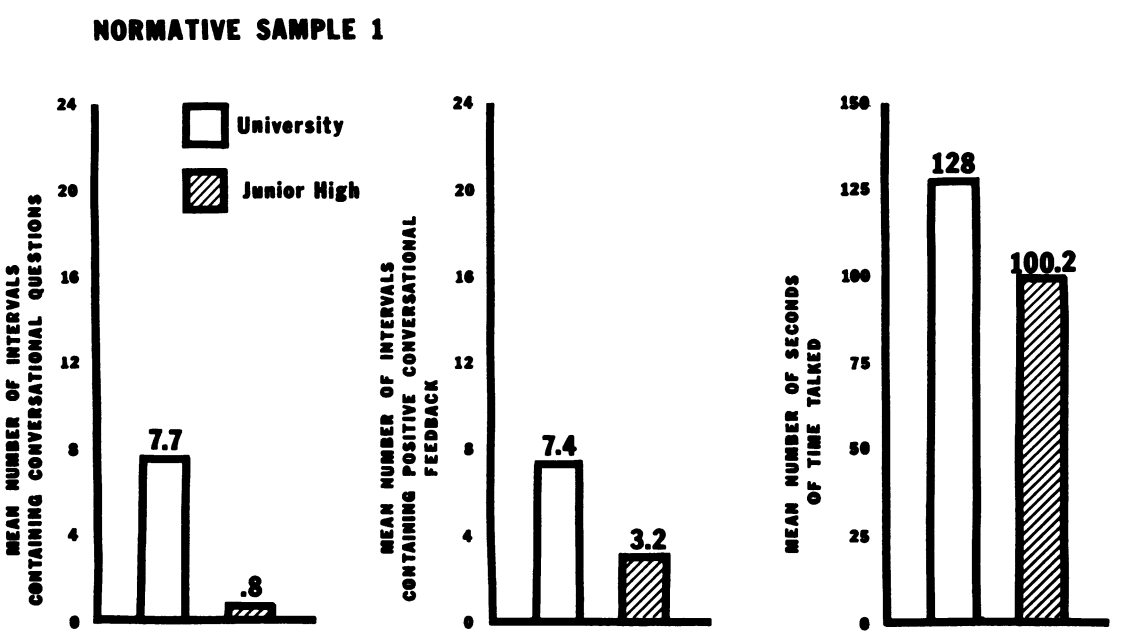


Fig. 1. Normative Sample 1. The left graph shows the mean number of 10-sec intervals that contained at least one conversational question asked by university females (left bar) and junior-high females (right bar) during their respective 4-min conversations with adults. The center graph shows the mean number of 10-sec intervals during the 4-min conversations that contained at least one instance of positive conversational feedback by the university females (left bar) and junior-high females (right bar). The right graph shows the mean number of seconds talked by the university and junior-high females during their respective conversations.

viewed and rated the randomized conversations in an order from one to 20. The second group viewed and rated the tapes in a counter-balanced order, first, conversations 11 through 20 were viewed, then conversations one through 10.

Before viewing the tapes, each judge was given a rating form and instructed to rate each conversant by making a mark along a seven-point bi-polar semantic differential scale (Osgood, Suci, and Tannenbaum, 1957) with the poles labelled "excellent" and "poor". The judges were also instructed to rate independently, and to try not to be influenced by the conversant's age or appearance. They were also told that they could go back and change any rating at any time. The judges received $4.00 each for their participation in the approximately 2.5-hr rating session.

A composite behavioral score was calculated for each conversant in each conversation in a manner that gave equal weight to a conversant's score on each of the three individual measures. While a score for conversational questions and positive conversational feedback could range from 0 to 24 (24, 10-sec intervals in 4 min), the time-talked score could range from 0 to 240 (240 sec in 4 min). Thus, to assure equal weighting, it was necessary to transform the time-talked scores. By dividing the total number of seconds talked by 10, the potential range of scores was also made 0 to 24. The composite score was computed by adding the conversational question and positive conversational feedback scores to the transformed score for time talked. For example, if a conversant had seven intervals in which she asked questions and seven intervals in which she emitted positive conversational feedback and spoke 103 sec, her composite score would be 24.3 (7 + 7 + 10.3).

*Results.* The mean rating by the judges of the university subjects was 5.25, with a rating of one being "poor" and a rating of seven being "excellent". The mean rating of the junior-high subjects was 3.4. The correlation between the

composite behavioral score and the judges' rating yielded a coefficient of 0.85. The correlation coefficients between each component behavior and the judges' ratings were 0.70 for conversational questions, 0.56 for positive conversational feedback, and 0.43 for time talked.

The 13 raters comprised different subgroups on the basis of sex, student-nonstudent status, and age. The degree of consensual conformity between these various subgroups was evaluated by correlating the average ratings of each rater subgroup (*e.g.,* females) with the average ratings of the corresponding subgroup (*e.g.,* males). The resulting correlations were 0.95 between the male ($n = 7$) and female ($n = 6$) subgroups, and 0.92 between the student ($n = 9$) and nonstudent ($n = 4$) subgroups. Further, the raters were divided into three age-related subgroups, (a) 21 and under ($n = 4$), (b) 22 to 34 ($n = 7$), and (c) 35 and older ($n = 2$). The correlation coefficient between groups a and b was 0.85, between a and c, 0.82, and between b and c, 0.95. The correlations all indicated a high degree of consensus among the various rater subgroups.

To determine the extent of interjudge agreement or reliability, the Kendall coefficient of concordance (W) was employed (Siegel, 1956). In using this test, rankings of all 20 conversations were obtained from each judge by rank ordering the conversations from highest rated to lowest rated by that judge. The resultant score of $W = 0.61$ was significant at the $<0.001$ level. This may be interpreted to mean that each of the judges applied essentially similar standards in evaluating the conversations.

To provide some estimate of intrajudge agreement (the agreement of a judge with himself), Pearson product moment correlations were computed between each judge's ratings of one of the conversations of each conversant and the remaining conversation of each conversant. The resulting correlations for each of the 13 judges were transformed into a single correlation coefficient using a Fisher Z transformation (Guilford, 1965). The resultant $r = 0.68$ suggests an overall consistency between the judge's ratings of individual conversants across their two conversations.

## Summary

The first part of this study indicated that some of the behavioral components of conversation could be specified and reliably measured. Data from the normative samples of conversation indicated that the university subjects used more of the specified components than the junior-high subjects. It also appeared that subjects who emitted more of the specified component behaviors were considered better conversationalists by adult members of the local community. Further, there appeared to be consensual agreement between the various raters and subgroups of raters as to relevant judgements of conversational ability. The 0.85 correlation coefficient between the composite behavioral score and the judges' ratings suggested a strong relationship between the specified behavioral components and how one was evaluated as a conversationalist. However, it was possible that the high correlation might have been a "chance" occurrence, due to the fact that the behavioral definitions were developed from the same tapes that were rated, and consequently the correlation might have been unique to that sample (Blumenfeld, 1972). Thus, a replication was necessary to verify the results.

## Sample 2

### *Subjects*

Five junior-high and five university females ranged in age from 12 to 20 yr. The junior-high students were in the eighth and ninth grades and the university students were freshmen and sophomores. As with the girls who participated in the first sample, the junior-high girls volunteered through an announcement made in a study hall and the university girls responded to an announcement made in a sorority house.

*Setting, Apparatus, and Procedures*

The setting, apparatus, and procedures were identical to those reported for Sample 1.

## RESULTS

The results indicated that the university subjects averaged 7.8 intervals in which they asked at least one conversational question, 8.3 intervals containing at least one instance of positive conversational feedback, and spoke an average of 137.6 sec. The junior-high subjects averaged 1.1 intervals in which they asked conversational questions, 3.6 intervals containing at least one instance of positive conversational feedback, and spoke an average of 113.4 sec.

*Social Validation*

*Judges.* Fifteen local community residents, seven males and eight females, volunteered to serve as judges. They ranged in age from 19 to 58 yr, average age, 27. This group of judges consisted of 10 graduate and undergraduate students, one student teacher, two homemakers, one bookkeeper, and one university housemother.

*Setting, apparatus, and procedures.* The setting, apparatus, and procedures were identical to those reported for Sample 1.

*Results.* The mean rating by the judges of the university subjects was 5.0, with one being "poor" and seven being "excellent". The mean rating for the junior-high subjects was 3.7. The correlation between the composite behavioral score and the judges' ratings was 0.84. The correlation coefficients between each component behavior and the judges' ratings was 0.63 for conversational questions, 0.64 for positive conversational feedback, and 0.65 for time talked.

Correlations to determine consensual conformity between the various subgroups yielded coefficients of 0.76 between the male (n = 7) and the female (n = 8) subgroups, 0.79 between the student (n = 10) and nonstudent (n = 5) subgroups. The raters were also divided

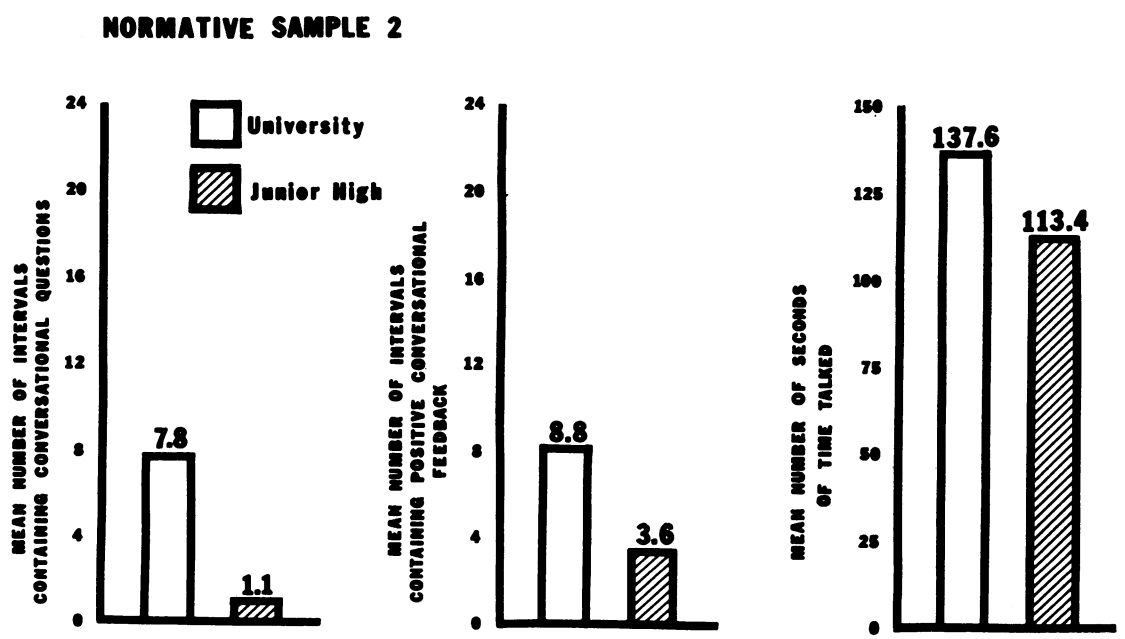


Fig. 2. Normative Sample 2. The left graph shows the mean number of 10-sec intervals that contained at least one conversational question asked by university females (left bar) and junior-high females (right bar) during their respective 4-min conversations with adults. The center graph shows the mean number of 10-sec intervals during the 4-min conversations that contained at least one instance of positive conversational feedback by the university females (left bar) and junior-high females (right bar). The right graph shows the mean number of seconds talked by the university and junior-high females during their respective conversations.

into three age-related subgroups (a) 21 and under (n = 8), (b) 22 to 34 (n = 4), and (c) 35 and older (n = 3). The correlation coefficient between groups a and b was 0.94, between groups a and c, 0.75, and between groups b and c, 0.78.

The Kendall coefficient of concordance (W) was again used to determine the extent of interjudge agreement. The resultant score of W = 0.46 was significant at the $<0.001$ level.

Pearson product moment correlations to provide an estimate of intrajudge agreement with himself were computed and transformed. The Fisher Z transformation yielded a coefficient of r = 0.61.

## II. TRAINING CONVERSATIONAL SKILLS

Achievement Place for Girls is a residential group home for six to eight court-adjudicated delinquent and predelinquent girls. Interested residents of the community make up the Board of Directors, who are responsible for establishing the major goals of the program. The goals are carried out by professional teaching-parents who live in the home and have primary responsibility for the treatment and care of the girls (Phillips, Phillips, Fixsen, and Wolf, 1974). One major goal of Achievement Place for Girls is teaching social-interaction behaviors that the community views as being important for successful relationships.

### *Subjects, Setting, and Apparatus*

Four girls in the Achievement Place program volunteered to participate in this aspect of the research. The girls ranged in age from 12 to 14 yr and were in the seventh through the ninth grades. The teaching-parents recommended asking these girls to participate on the basis that they were "generally deficient" in social communication skills with adults.

The setting and apparatus were identical to those in which Samples 1 and 2 were obtained.

### *Procedures*

A multiple-baseline design (Baer *et al.*, 1968) across the behaviors of conversational questions and positive conversational feedback was used to analyze effectiveness of the training procedures. Each subject participated in three to six 4-min baseline conversations with previously unknown adults. The conversants received the same instructions as those participating in collection of the normative data. The baseline conversations involved two after-school sessions of approximately 1.5 hr and each girl was paid $1.00 before the session.

The procedure for training conversational questions consisted of three parts: instructions with rationale, demonstration, and practice with feedback. Instructions with rationale consisted of describing the behavior, giving oral and written examples of the behavior, explaining the importance of the behavior in conversation, and asking the girls why they felt the behavior was important. Demonstration consisted of two experimenters modelling the behavior in a sample two-person conversation. In practice with feedback, 4-min interactions with one of the experimenters were videotaped to allow feedback to each girl on the amount of the target behaviors she had engaged in. When the girls met the experimenter-established criterion (16 instances of conversational questions in each of two consecutive 4-min conversations with one of the experimenters) the girls were asked to participate in additional conversations with unknown adults, earning $0.10 for each conversational question they would ask.

When the girls had demonstrated proficiency in asking conversational questions, they were taught to give positive conversational feedback. The procedure for teaching positive conversational feedback was identical to that employed in the teaching of conversational questions: instructions with rationale, demonstration, and practice with feedback. On meeting the experimenter-established criterion of 16 instances of positive conversational feedback in two con-

secutive 4-min conversations with one of the experimenters, the girls were informed of the opportunity to engage in more conversations with unknown adults. They each continued to receive $0.10 for each instance of positive conversational feedback.

One girl, Kim, did not participate in the group training session. Kim was trained in part by viewing a videotape of the instructions and demonstrations provided in the group training session. The practice and feedback components of the training were identical to the group sessions. Again, a multiple-baseline design was used across Kim's behavior.

## RESULTS

In baseline, Kim asked questions in an average of three of the 24 intervals and in each conversation gave positive conversational feedback in an average of two intervals. In the posttraining sessions, she asked questions in an average of 13 intervals and gave positive feedback in an average of 11 intervals.

Mary did not ask any questions in her baseline conversations and had an average of 0.8 intervals in which she gave positive feedback. In her posttraining sessions, she asked questions in an average of 15 intervals and averaged 17 intervals of positive feedback.

Lynn, in her baseline sessions, asked questions in an average of five of the 24 intervals and gave positive feedback in an average of 1.7 intervals. In her posttraining sessions, she averaged 18 intervals in which she asked questions and averaged 17 intervals in which she gave positive feedback.

Diane averaged 0.5 intervals containing questions and averaged 9.2 intervals containing positive feedback in baseline. In posttraining sessions, Diane asked questions in an average of 12 intervals and gave positive feedback in an average of 18 of the 24 intervals.

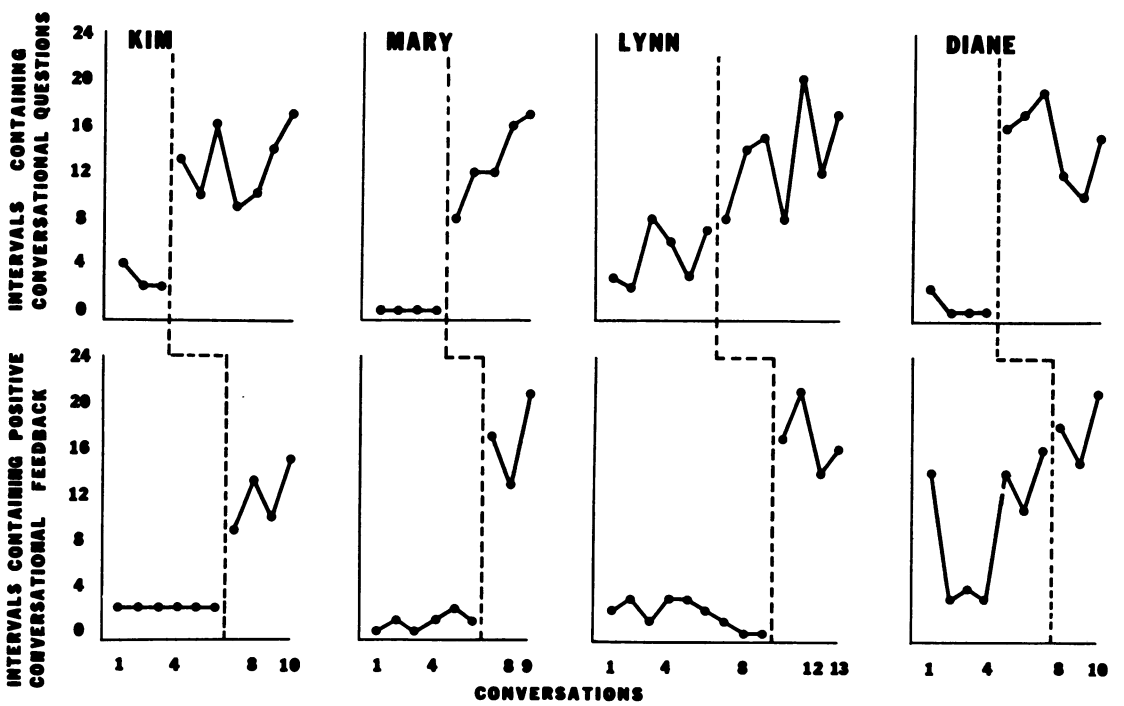


Fig. 3. The number of 10-sec intervals that contained at least one conversational question (upper graphs) and the number of 10-sec intervals that contained at least one instance of positive conversational feedback during conversations with previously unknown adults for the four Achievement Place girls. The vertical dashed line represents where training was introduced.

SUMMARY

Training effectively increased the use of conversational behaviors. After training, the Achievement Place girls consistently asked more questions and gave more positive feedback than the "normal" junior high-school girls.

The experimental design demonstrated that the effect could be replicated across behaviors and across subjects. On the other hand, the roles of the various components of the training package were not analyzed. However, the use of instructions and rationale, demonstration, practice, and feedback coupled with motivation has proved to be practical and effective in demonstrating behavior change with delinquent and predelinquent youth (Braukmann, Maloney, Fixsen, Phillips, and Wolf, 1974; Werner, Minkin, Minkin, Fixsen, Phillips, and Wolf, 1975). While the training package had proven effective in increasing the behavior of the girls, a most important question remained. Could the effects of the training be validated as producing "socially meaningful" change (Serber, 1972)?

## III. SOCIAL VALIDATION OF BENEFITS OF CHANGE

Although the Achievement Place girls demonstrated proficiency in using the specified and validated behaviors, the qualitative effects of the training remained unknown. Thus, an attempt was made to determine if the girls were viewed as better conversationalists after training.

### *Subject, Setting, and Apparatus*

The 15 adults who viewed and rated the second set of normative samples served as judges. The setting and apparatus were identical to those used in the rating procedures of the second set of normative samples.

### *Procedures*

One baseline and one posttraining videotaped conversation for each Achievement Place girl were randomly selected for judging. The eight conversations were then randomly ordered, with the provision that no subject would appear in two consecutive conversational sequences.

Immediately after rating the 20 conversations constituting the second set of normative samples, the judges viewed and rated the baseline and posttraining conversations of the Achievement Place girls. The first group of judges viewed and rated each conversation in an order from one to eight. The second group viewed and rated the tapes in a counter-balanced order, *i.e.,* first conversations five through eight, then conversations one through four. The judges were not informed that Achievement Place girls were involved in the last eight conversations and no distinction was made between the 20 normative sample tapes and the last eight videotapes.

### RESULTS

Figure 4 shows that the average rating of the Achievement Place girls before training was 2.8 and the average rating of their conversational ability after training rose to 4.3.

Figure 5 shows that the individual rating for each Achievement Place girl increased after training. Before training, Diane received an average rating of 2.1, which rose to 4.2 after training. Fourteen of the 15 judges rated Diane as a better conversationalist; one judge's rating remained the same. Before training, Mary received an average rating of 2.7 and after training it rose to 3.5. Twelve judges rated Mary as a better conversationalist, two rated her ability as the same, and one rated her ability to have decreased. Lynn averaged 2.8 before training and 4.7 after training. Eleven judges rated improvement and four judges' ratings remained the same. Kim's average rating before training was 4.1 and after training was 4.7. The rating change seen for Kim was less dramatic than for the other girls. Nevertheless, seven judges rated improvement, five judges' ratings remained the same, and three judges rated her ability to have decreased.

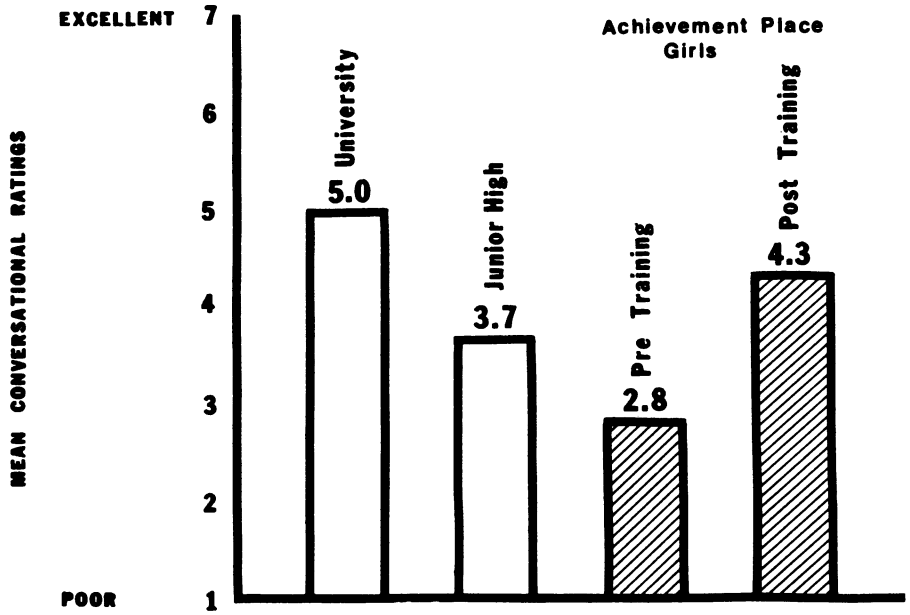


Fig. 4. The four bars respectively represent the mean ratings of the conversational ability of the university females, junior-high females, Achievement Place females before training and the Achievement Place females after introduction of the training package, as judged by residents of their local community using a seven-point bi-polar scale with the poles labelled *poor* and *excellent.*

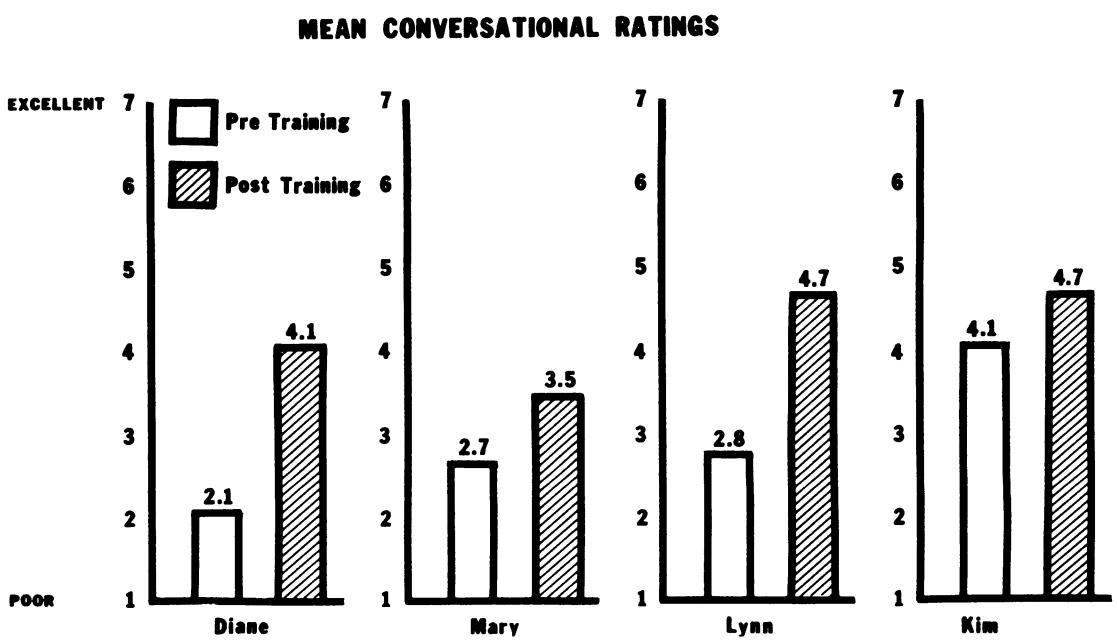


Fig. 5. The mean conversational ratings of each of the four Achievement Place females before and after introduction of the training package as judged by residents of their local community using a seven-point bi-polar scale with the poles labelled *poor* and *excellent.*

## GENERAL DISCUSSION

The results suggest that some of the behavioral components of conversation can be reliably specified and socially validated as being important. The results also suggest that the specified components can be trained and the benefits of increases in the behavior of the trainees can also be socially validated.

Videotaped conversations of "normal" junior-high and university females provided normative information about their conversational behaviors. The older females generally asked more questions, gave more positive feedback, and spoke more. Reliable definitions of these behaviors were developed, but their relational importance to conversation remained unknown. Thus, to quantify conversation, a social-validation procedure, which consisted of ratings, by relevant judges, was used to establish the social importance of these behaviors.

Once validated, the behaviors were taught to girls who appeared to be behaviorally deficient in speaking with adults using a multiple-baseline design across behaviors. The results indicated that all subjects increased their usage of the specified conversational components. However, the question still remained as to whether the girls would be considered better conversationalists by adult members of their local community. To provide an answer, adult judges from the girls' local community evaluated their conversational skills by rating videotaped conversations before and after training. The results, showing increased skill levels, socially validated the benefits of the behavioral increases.

Measurement in applied behavior analysis has traditionally been concerned with the measurement of objectively defined behaviors, *i.e.,* behaviors that could be reliably recorded as to their presence or absence by independent observers. As behavior analysis expands into more complex realms of behavior, the social validity of objectively measured behaviors will become a more significant issue. As we attempt to deal with more complex behaviors we may inadvertently choose to measure behavioral components that are not relevant to the complex behavior of interest to society. Thus, our traditional objective measurement procedures will have to be supplemented by methods to establish the relationship of the specified objective components to the complex class of behavior. Social validation is one such method.

For example, affection might be considered a complex social behavior. If the goal of a behavior analyst is to teach a parent to be more affectionate towards his child, it might be necessary to specify the important component behaviors of affection. Some of the components might include touching, smiling, and hugging. To validate the social importance of these behaviors, four steps might be used. First, gathering sample parent-child interactions. Second, developing and recording reliable definitions of specific interactant behaviors. Third, employing relevant judges, *e.g.,* other parents, randomly selected if possible, to rate the sample interactions and evaluate each parent as to the amount of affection shown to the child within the interaction. The evaluation instrument might be a bi-polar rating scale with the poles labelled as to the amount of affection shown. Step four would involve correlating the ratings of the judges with a composite score of the objectively measured behaviors of the parents. The subsequent coefficient would indicate the level of relationship of the specified objectively measured components of affection to the common English "meaning" of affection as rated by the judges. In the present study, the use of a composite behavioral score produced a measure highly correlated with the ratings.[3] Some of the important behavioral components of creativity, conversation, and affection, as well as other complex classes of social behaviors, could probably be identified through the use of social-validation procedures.

[3]Each of the components was given an equal weighting in the composite score in this study. There are more statistically sophisticated procedures for assigning relative weights that involve regression equations.

Social-validation procedures can also be used to identify and describe subsequent benefits of a behavior change. Traditionally, behavior analysts have evaluated the effectiveness of their training procedures almost exclusively in terms of their ability to modify objectively defined behaviors. However, training procedures that increase objectively defined behaviors may not necessarily produce an increase in the perceived skill of the subject according to relevant judges. The teaching of socially validated behaviors does not necessarily produce socially valid changes. Again, evaluations by relevant judges could be a means of verifying the benefits of behavior change. Evaluations of a subject both before and after training would provide the behavior analyst with information as to the effectiveness of his intervention procedures.

Since the present research was conducted in a structured training setting, rather than in the natural environment, the degree to which equally beneficial changes in conversational skills can be produced in the natural setting is a question for future research. Future research should also examine a wider range of relevant judges, including peers and the youths themselves.

## REFERENCE NOTE

1. Wolf, M. M. and Sherman, J. A. *The principles and procedures of behavior modification.* In D. Hursh, J. Wildgen, B. Minkin, N. Minkin, J. Sherman, and M. Wolf (Eds.), Lawrence, Kansas: Project MORE, 1973.

## REFERENCES

Arrington, R. E. Time sampling in studies of social behavior: A critical review of techniques and results with research suggestions. *Psychological Bulletin,* 1943, **40,** 81-124.

Baer, D. M., Wolf, M. M., and Risley, T. R. Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis,* 1968, **1,** 91-97.

Bijou, S. W., Peterson, R. F., and Ault, M. H. A method to integrate descriptive and experimental field studies at the level of data and empirical concepts. *Journal of Applied Behavior Analysis,* 1968, **1,** 175-191.

Blumenfeld, W. S. I am never startled by a fish. *American Psychological Association Monitor,* 1972, **3,** 14.

Braukmann, C. J., Maloney, D. M., Fixsen, D. L., Phillips, E. L., and Wolf, M. M. An analysis of a selection interview training package for predelinquents at Achievement Place. *Criminal Justice and Behavior,* 1974, **1,** 30-42.

Coleman, J. C. *Abnormal psychology and modern life.* 3rd ed.; New York: Scott Foresman & Co., 1964.

Felixbrod, J. J. and O'Leary, K. D. Effects of reinforcement of children's academic behavior as a function of self-determinant and externally imposed contingencies. *Journal of Applied Behavior Analysis,* 1973, **6,** 241-250.

Guilford, J. P. *Fundamental statistics in psychology and education.* 4th ed.; New York: McGraw-Hill, 1965.

Haase, R. F. and Tepper, D. T. Nonverbal components of empathetic communication. *Journal of Counseling Psychology,* 1972, **19,** 417-424.

Harris, F. R., Wolf, M. M., and Baer, D. M. Effects of adult social reinforcement on child behavior. *Young Children,* 1964, **20,** 8-17.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. *The measurement of meaning.* Urbana: University of Illinois Press, 1957.

Phillips, E. L., Phillips, E. A., Fixsen, D. L., and Wolf, M. M. *The teaching-family handbook.* Lawrence: University of Kansas Press, 1974.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Serber, M. Teaching the nonverbal components of assertive training. *Journal of Behavior Therapy and Experimental Psychiatry,* 1972, **3,** 179-183.

Smith, H. C. *Sensitivity training: the scientific understanding of individuals.* New York: McGraw-Hill, 1973.

Truax, C. B. and Carkhuff, R. R. *Toward effective counseling in psychotherapy.* Chicago: Aldine, 1967.

Werner, J. S., Minkin, N., Minkin, B. L., Fixsen, D. L., Phillips, E. L., and Wolf, M. M. Intervention Package: Analysis to prepare juvenile delinquents for encounters with police officers. *Criminal Justice and Behavior,* 1975, **2,** 55-83.

*Received 11 November 1974.*
*(Final acceptance 23 July 1975.)*