

#19

~~X~~  
EVALUATION OF A PROGRAM TRAINING POLICE IN INTERPERSONAL  
COMMUNICATION SKILLS

Edwin Susskind, University of Maryland Baltimore  
County

Derek Roemer, Southwestern Community Mental Health  
Center, Baltimore

Joel M. Lazar, Spring Grove Hospital Center

Oresto R. Digiondomenico, University of Maryland  
Baltimore County

Steven L. Wise, University of Maryland Baltimore  
County

presented to

The National Conference on Criminal Justice Evaluation

Washington, D. C.

February 23, 1977

54788

This project is the product of a collaboration between the Baltimore County Police Department, the Baltimore County Health Department, the Maryland Governor's Commission on Law Enforcement and the Administration of Justice, and the graduate program in community-clinical psychology at the University of Maryland Baltimore County. We would like to express our appreciation to: Chief Joseph Gallen and former-chief William Ensor of the B.C.P.D.; Kenneth Hines, Carl Banaszewski and Tom Saleme of the Governor's Commission; Connie Caplan, Criminal Justice Coordinator for Baltimore County; Mehdi L. Yeganeh, M.D., M.P.H., Director of the Baltimore County Mental Health Bureau, and Sidney Wolf, Ph.D., formerly of that bureau; Donald Bartnick and Jim Koch, graduate students at U.M.B.C.; Wendell Williams, psychology intern at Spring Grove Hospital Center.

I would like today to present an overview of the evaluation strategies used to assess a program which trained Baltimore County police officers in interpersonal communication skills. My primary goal is to share our deliberations concerning dependent measures and design. In addition, I will report the results of our first year evaluation. Hopefully, this discussion will be of use to other evaluators, as well as a source of input and suggestions for our evaluation team.

#### Description of the Training Program

Let's begin with a brief description of the training program and its rationale. The program attempts to broadly influence police officer's patterns of communication with non-criminal citizens. Some examples of non-criminal citizens would include: (1) a citizen reporting the theft of his car; (2) a parent who has just learned that his teenager has been arrested or has run away; (3) a couple having a marital dispute. As Helena Carlson pointed out in her presentation to this conference yesterday (Carlson and Sutton, 1977), the provision of "help-providing interpersonal services" to non-criminal citizens is the major function of police work. Similarly, Epstein (1962) has estimated that this function accounts for 90% of a typical police officer's time. Further, these contacts do much to

determine citizen attitudes towards local police. For example, Carlson and Sutton found that a victimized citizen's satisfaction with police services had less to do with whether the crime was solved than with the quality of their interpersonal interaction with the officer.

As we have become more aware of the importance and job-relevance of an officer's interpersonal communication skills, one response has been to develop training programs for police "specialists" in crisis intervention. The pioneering work in this area has been done by Bard and his colleagues (Bard and Berkowitz, 1967; Zacker, Rutter and Bard, 1971; Bard and Zacker, 1976). In contrast to such "specialist" programs, our goal was to provide a training package in basic interpersonal communications that would reach nearly all patrol officers in the Baltimore County Police Department.

The history of our program begins in the spring of 1975, when the Baltimore County Police Department asked the county Health Department to provide an officer training program in interpersonal communication. The Governor's Commission on Law Enforcement and the Administration of Justice supported the project. The health department then developed a model which drew on Carkhuff and Truax's work in training counselors. Each training module exposes a group of approximately 14 officers to about 60 hours of training, spread over six weeks. During the module, officers discuss and role play examples of their contacts with citizens.

During the first year of the program a total of five modules were conducted, in which a total of 73 officers participated. Recruitment of participants started at the higher levels of command in the patrol division, namely, police captains. Subsequently we recruited lieutenants, sargeants, and a few patrolmen. During the second year of the program, we will focus on officers below the rank of sargeant.

#### Evaluation Criteria

How does one go about evaluating such a program. I would like to review 7 criteria or questions that we have considered.

##### (1) Administrative Criteria

Here the basic question is: did the program train the number of officers it had contracted for? Were the operational mechanics successfully arranged? Were there problems of absenteeism, attrition, or refusal to participate in the program?

By these criteria, our program was relatively successful. A total of 73 officers were trained in 5 modules. This does fall short of the original projection of 110 officers to be trained in 6 modules. However, one must consider that the program had to cope with a major, unanticipated reorganization and reassignment of command officials in the police department. The program did gain the support of the police department: there were no problems of absenteeism or attrition; the department agreed to continue the program

for a second year and to absorb the cost of compensatory time for the officers in training.

The program passed its first hurdle, in that it brought officers to the training sessions. But, would the officers learn anything in these sessions? This leads us to our second criterion.

(2) Test of Skill Acquisition

Here our basic question is: Did the participants learn the concepts and skills that we sought to teach?

To answer this question, we made use of a video-taped test, The Counseling Skills Evaluation, developed by Dr. Sidney Wolf (1976). Officers were presented with a video-taped vignette in which an actor role plays a person who is presenting a problem and seeking help. After each vignette, the officer is asked to write the response he would make if faced with such an individual. The adequacy of this response is rated on a 5 point rating scale, and we refer to it as the officer's communication score. After he has given his own response, the officer is shown five different video-taped "helper" responses. He is asked to rate the helpfulness of each of these five responses. His ratings are compared with a set of standard ratings made by experts. This comparison yields a discrimination score, which measures the officers ability to discriminate the most helpful response.

The test includes a total of 12 vignettes. The scores are summed across the vignettes, to yield two measures for each officer: a communication score and a discrimination score.

For purposes of our evaluation, officers were given six of the vignettes as pre-test at the beginning of training. The remaining six vignettes served as a post-test after training. This procedure was followed for all modules except module 1.

The vignette discrimination data for module number one were analyzed to examine any differences in difficulty between the vignettes. Module number one was unique in that all 12 vignettes were presented at one time. When the first six vignettes (normally comprising the pre-test) were contrasted with the last six vignettes (normally the post-test), the analysis of variance found no significant difference for the discrimination scores. There was a significant difference in the communication scores; the post-test items appear to be somewhat more difficult than the pre-test. Thus, to the extent that there was a systematic difference between the pre and post vignettes, the test was biased against the training program, rather than in favor of the program.

An analysis of variance was performed on the discrimination scores for fifty-two officers in modules 2 - 5. There was a highly significant improvement from pre-test to post-test ( $F = 275.25$ ,  $df = 1, 48$ ,  $p < .0001$ ). Further, this improvement was significant for each of the modules, when analyzed separately.

These discrimination data indicate that officers improved their ability to recognize appropriate responses.

But, did they also improve their ability to produce such responses? An answer is provided by the communication score data. We have pre and post training communication scores for modules 3 - 5, for a total of 41 officers.

The communication scores for these officers were computed independently by two experienced raters. The inter-rater reliability coefficients were high (.78 for the pre-measures and .95 for the post-measures). An analysis of variance showed that the post-scores were significantly better than the pre-scores ( $F = 31.42$ ,  $df = 1, 38$ ,  $p < .0001$ ). We conclude that not only did the officers learn to identify helpful responses, but they were able to generate them spontaneously.

The vignette test data allow us to conclude that we have satisfied the second of the criteria in our list. That is, the officers did learn the concepts we sought to impart.

Having learned these concepts, did the officers find them applicable to their work situations? This is our third criterion.

(3) Officer Perception of the Contribution of the Training to their Job Performance.

Nearly all (69/73) of the officers participated in a one hour interview, which was administered three to six months after the officer had completed training. Part 1 of the interview consisted of ten general, open-ended, non-directive questions. For example: Do you feel you have changed as a result of this program? If so, how? These



questions were intended to produce relatively "spontaneous" "non-prompted" responses. In contrast, Part 2 of the interview included eighteen highly-specific multiple choice items. Each question essentially had the format "Are you better able to ..... as a result of the training?" For example: "Do you feel that you are better at handling other people's anger as a result of the training?" Officers chose an answer from a five point ordinal scale that ranged from " 1 = not at all" to " 5 = yes, very much improved." Officers were assured that their responses would be kept anonymous.

How does one make use of such 'soft' interview data in a manner that commands some face validity? We are all aware of the effects of response biases and demand characteristics on interview data.

One measure that we employed was the number of officers who, in Part 1, cited some specific positive change in themselves which they attributed to the training. This percentage ranged from a low of 54% to a high of 88%, across the five modules.

We have reason to believe that these percentages have face validity. It is not a trivial task to get a group of police officers to acknowledge that they have benefitted from a "psychological" training program. Further, as the percentages indicate, there were officers in each module who felt free to state the opinion that the program was worthless. Overall, however, a majority of the officers stated that the training had helped them.

A second measure which lent face validity to the interview was the rich anecdotal data spontaneously offered in response to the non-directive questions. That is, officers did not merely state that the program had been helpful; rather, they provided convincing, detailed and at points dramatic examples. Consider the following two examples:

(a) I can listen closely. Let others tell their views and try to solve problems before you help. There was a child abuse case. The mother-in-law was an alcoholic. Before the training I might not have listened to her .... would not have trusted anything she said. The story was unbelievable. Because we listened, we got a 7 year old child out of the closet that had been there for a year.

(b) A woman came in the other day, all worked up and nothing that I would have said could have calmed her down. Before, I might have become very annoyed. I just let her have her say and then she left. She just had to get it off her chest.

We have included a set of such anecdotes in our annual evaluation report. On the one hand, they are useful as measures of outcome. In addition, they are important process measures; these process measures can help the trainers to specify the relevance of the training to the daily work of policemen.

A third measure was drawn from the interview, based on the 18 multiple choice items in part 2. The 18 items were divided into four Clusters. Cluster I concerned the development of specific communication skills, which were the goal of the training program. If the program had in fact helped the trainees and if the interview responses did not simply reflect a halo effect, then the items in Group I should re-

ceive relatively high scores. In contrast the items in Cluster II, which concerned "dealing with others feelings," might be expected to receive somewhat lower scores. Similarly the items in Cluster III, which concerned "dealing with ones own feelings," might be relatively lower. Finally, we would expect the items in Cluster IV to be lowest of all. Cluster IV concerned changes in the officer's attitude toward the community and toward various sub-groups in the community, rather than toward individuals. Change of this sort was not an explicit goal of the training.

The outcomes were fully consistent with our prediction. In each of the five modules, the Cluster I items received the highest overall mean rating, followed respectively by Clusters II, III, and IV. The overall mean for Cluster I was 3.21, on a five point rating scale where "3" means "yes, a little improved."

We conclude that: (1) a majority of the officers found the program helpful; (2) they found it most helpful specifically in the area of communication skills; (3) this differential positive rating of the communication skill items lends support to the validity of the questionnaire.

Having ascertained that officers perceive a job-relevant improvement in their own communication skills, we might next logically ask "are the trained officers perceived more favorably by citizens?" This leads to our fourth criterion.

(4) Citizen Perception of Police Officers (an interview)

We are currently planning a brief telephone interview with non-criminal citizens who have had contact with police

officers. From the incidence reports filed by each officer, we will locate such citizens who were contacted by the officer prior to his training and others who were contacted after training. In addition, we will have a control group of citizens who were contacted by officers who never underwent training. Note: this will restrict our sample to citizens who own private telephones.

The usefulness of such telephone interviews has been criticized recently by Bard and Zacker (1976). However, a number of researchers favorably report their use as measures of police performance (Driscoll et. al., 1973; Carlson and Sutton, 1977; Snortum and Pearce, 1977).

One can go beyond the measurement of officer and citizen perceptions by directly observing their behavior and interaction. This leads us to a fifth criterion measure.

(5) Interaction Analysis of Police and Citizen Behavior

There is no question but that such direct observation of behavior would augment and possibly surpass our interview data. However, we decided to forego this approach for three reasons:

- (a) the data are expensive to collect. For example, it is difficult, perhaps impossible, to "schedule" such police-citizen contacts in advance.
- (b) the procedures are highly obtrusive and may influence the behavior of "the actors."
- (c) the police officers may resent the procedures as being intrusive.

However, there is another hard-data alternative to direct observation, namely, the relatively non-reactive, archival data collected by the police department. Such data comprise our sixth criterion measure.

(6) Archival Data

Bard and Zacker (1976) provide an example of outcome measures drawn from police archives. Their officers were trained to intervene in domestic disputes. A major finding in Bard's evaluation was a decrease in the number of assaults and casualties suffered by officers after training. Similarly, Bard examined "days of sick leave" as a measure of job stress.

We have been in the process of identifying additional relevant variables which can be culled from police records. For example, we are interested in "nuisance arrests", that is arrests which in some cases might have been avoided had there been better communication between officer and citizen. For example, consider arrests for: disturbing the peace; disorderly conduct; pedestrian drunkenness; loitering; domestic disputes; failure to obey an officer; insulting an officer; resisting arrest.

Such measures are economical and relatively non-reactive. They have high face validity for both researchers and police administrators as measures of job performance. However, they pose some problems. The police department or union may be reluctant to release such information to "civilians". Departmental computers may have such data available only

by geographic units (districts, sectors, reporting areas) rather than by individual officer.

There is a further, logical problem. The officers job performance is a function of many factors that have nothing to do with the training program: for example, the ethnic and economic characteristics of his 'beat', the weather and season of the year, the state of the national economy. It may be that the communications skill training significantly improves his interpersonal skills, without its achieving a statistically significant change in the archival measures we've suggested. That is, these archival measures may be relatively insensitive.

In contrast, one can argue that our second criterion measure, the video-taped test of communication skills, will be the most sensitive, since it should be influenced primarily by the training and not by other uncontrolled variables. In fact, one would make the following prediction: If a measure of effect magnitude (like omega-squared) were calculated for each of the criterion measures we have discussed, then the effect would be found to be largest for the test of communication skills, next largest for the officer interview, followed respectively by the observations of officer behavior, the behavior and interviews of citizens, and finally by the archival measures. In other words, the statistical sensitivity of these outcome measures should vary directly with their logical and temporal proximity to the training program.

In closing, I would like to indicate a seventh criterion measure, which may be untraditional or debatable, and which has received insufficient attention.

(7) Dissemination of Information

I propose that one measure of the utility and merit of a program is not only the service it provides to its specific trainees, but rather the service it provides to the profession and the community at large. For example, our training project has produced an extensive, useful annotated bibliography which was included in the evaluation report. Similarly, the dissemination of information through conferences, like today's, and through publications should be valued by the program's evaluators and funding agencies.

References

- Bard, M. and Berkowitz, B. Training police as specialists in family crisis intervention: A Community Psychology Action Program. Community Mental Health Journal, 1967, 3, 315-317.
- Bard, M. and Zacker, J. The Police and interpersonal conflict: third party intervention approaches. Washington, D. C.: Police Foundation, 1976.
- Carlson, H. and Sutton, M. Consumer Evaluation of Police Performance. Paper presented to the National Conference on Criminal Justice Evaluation, Washington, D. C., February 22, 1977.
- Driscoll, J., Myer, R. and Schanie, C. Training police in family crisis intervention. Journal of Applied Behavioral Science, 1973, 9(1), 62-82.
- Epstein, C. Intergroup relations for police officers. Baltimore: Williams and Wilkins, 1962.
- Snortum, J. and Pearce, J. Police effectiveness in handling disturbance calls: An Evaluation of crisis intervention training. Paper presented to the National Conference on Criminal Justice Evaluation, Washington, D. C., February 23, 1977.
- Wolf, S. and Wolf, C. The counseling skills evaluation manual. Towson, Md.: The Psychological Skills Development Corporation, 1976.
- Zacker, J., Rutter E. and Bard, M. Evaluation of attitudinal changes in a program of community consultation. Community Mental Health Journal, 1971, 7, 236-241.