

Evaluate Your Training Program

LESTER TARNOPOL
Luckmann-Tarnopol & Associates
Management Consultants
City College of San Francisco

"We don't have to defend training! Evaluation of training programs is unnecessary." This is a statement we sometimes hear from training directors about human relations training. "Besides, you can't evaluate such an intangible thing as human relations," is another rationalization often used. True, evaluation of human relations training presents some difficulties, but these are not insurmountable. Evaluation depends on measurement, and human behavior can be measured.

THERE are several very important reasons for evaluating your training program. First, a good evaluation instrument may often be used to tell you what kind of training is most needed by the supervisors. It will bring into focus both the strong points and the weak points of your supervisors. Second, this instrument may be utilized as a major source of factual information about supervision, communications, morale, and employee relations. And thus the instrument itself can constitute the subject matter for a very realistic training program. Third, the evaluation will tell you how much progress the supervisors are making back on-the-job where it counts. Fourth, evaluation will tell you where your training program has succeeded and failed and will help you improve your program. Finally, evalu-

ation will tell you which supervisors need further training and in what specific areas training is indicated.

To be of any real worth, both the evaluation instrument and method should be thoroughly scientific. We certainly don't wish to fool ourselves. And if the approach used fails to meet the requirements of scientific evaluation, we have NO way of knowing what results have been achieved. Happily, it is possible to plan and adhere to a scientific scheme of appraisal. We believe that the Employee Attitude Survey is the easiest, valid approach to evaluation of training programs. We shall therefore discuss some of the features of the surveys which are of importance to training directors.

LET us consider the measuring instrument. The supervisors are to be trained in human relations, therefore we wish to measure their human-relations behavior. The best way to approach this problem is to measure their behavior directly. This will minimize various types of errors and related problems. It is the supervisors' interpersonal relations with their employees on-the-job that we are really interested in. Therefore, this is what we should measure. The simplest, most direct manner of accomplishing this is through the Employee Attitude Survey. By asking questions re-

lated to all aspects of supervision, morale, communications, participation, etc., we can ascertain the behavior of each individual supervisor who has some minimum number of employees reporting to him. In our experience, five employees is a good minimum. Most supervisors will have between five and twenty-five people who report directly to them and who are therefore in a position to answer questions factually about them. By summing the employees' questionnaire responses for each supervisor, we have an adequate measure of each supervisor's behavior.

NO doubt you can foresee a number of possible sources of error when measuring supervisors' behavior by a survey. These possible loop-holes should be anticipated and reduced. For example, some employees will tend to be very suspicious of the survey and will fear giving truthful answers which they believe could be used against them. From our experiences, we feel that these suspicions probably can never be totally allayed. However, we will present evidence that very adequate quantitative results can be attained by following a few logical precautions. These obviously involve gaining the confidence of the employees. Here are some examples of the extent to which such confidence may be secured.

A supervisor offered the following comment on his survey questionnaire. "Regarding my relations with my supervisor, I will state we are having more meetings, but as for group decision and action we are about the same as when this training started. Responsibility is still not delegated. Communications give

us only a smattering of what I feel we should have as supervisors to do a good job. Part of our problem I am sure is the relationship of my supervisor with his boss. Change is taking place but it is slow and I know I am impatient at times."

THIS is indeed a candid statement, especially in view of the fact that it was necessary to identify all supervisors on this survey. This man is obviously saying that these are his real problems. Moreover, he is expressing confidence that his boss will not be informed. And even more important, he expects that this information will be used in the training program to induce corrective action.

A number of employees in a small work group contributed statements, all in a similar vein, on their survey questionnaires. Although their questionnaires were not signed, individual identification could have been possible by means of their background data and handwriting in such a small group. Nevertheless, they were unusually frank. Here is a sample statement.

"My supervisor won't take a definite position on matters. He's very nervous and conveys his nervousness to his subordinates. He can't make decisions. And even on small decisions he has to see the department head before he'll tell you what to do. He's afraid to do anything on his own and in his own words has often said, 'I'm not going to stick my neck out.' In other words he hasn't any backbone. The result is that to get any work done you have to make decisions for him. He doesn't take his employees' feelings into account in his re-

lations with them and therefore makes, in my opinion, a very poor supervisor."

HERE again the employee clearly has confidence in the anonymity of his survey replies. Moreover, he is not merely blowing off steam, he expects corrective action to be taken. The workers in this department later discussed these problems with us in individual interviews. They described their plight in great detail and expressed confidence that we would try to help them.

Such rapport with employees and supervisors can be built only through a long term relationship. In this case, we are discussing the results of four years of contact, and our second survey together. Although on the first survey many statements were written in, they were not quite so frank. However, the replies to the individual questions on *both* surveys were found to be extremely reliable by two different measures which we will discuss later.

Several precautions were taken in order to promote both the reliability and validity of the survey data. First, we spent a year getting acquainted with the personnel of the company before holding a survey. Second, the survey questions were forged by a committee of employees and supervisors cross-sectioning the organization, together with the consultants. Third, the entire company was involved in the survey, because the committee solicited ideas for survey questions from their fellow workers over a three month period. Fourth, multiple choice questions were used so that the employees only had to check their replies. Fifth, the questionnaires were not signed. Sixth, background information

was relegated to the last page of the questionnaire. Seventh, the employees were invited to discuss the coming survey in meetings with their supervisors. Eighth, they were told in these meetings that their questionnaires would be seen only by the consultants, who would prepare the information by work groups so that individual anonymity would be preserved. Finally, the survey questions were pretested before they were used in the actual survey.

After all of these precautions, some of the workers still remained suspicious. In a few cases, we learned that men checked themselves as "female" in order to throw snoopers off the track. However, experience in general indicates that once employees start answering a questionnaire they tend to reply according to their real attitudes and feelings. This can be proved by reliability measurements.

IT takes considerable professional experience to construct a good survey questionnaire. So it is advisable to have such assistance available before starting. Canned questionnaires are available, but they are not likely to meet your needs. To accurately measure the results of a training program, you should construct a questionnaire specifically suited to the requirements of both your company and your training program.

Involvement of the entire company by means of a survey committee cross-sectioning your company will pay dividends in ego-involvement. This survey committee can solicit ideas for questions from all the workers. The employees will contribute a number of new ideas which might otherwise be lost. The

professional researchers can then help the committee develop suitable questions. Finally, the committee can be of value in the pretesting of these questions.

IT is essential to have questions which reflect the state of supervision both directly and indirectly. As an illustration, our survey tapped the following areas: supervisory practices, employee morale, employee participation in meetings with supervisors to solve mutual problems, two way communication, employee knowledge of company practices, job information available to employees, employee work adjustment, working conditions, change in morale, supervision, and background information. Each of these areas was covered by from eight to thirty questions depending upon the number of subdivisions in the area. All of this must be carefully figured out before completing the survey questionnaire, so that each area and subarea will be adequately represented by questions.

This type of survey may contain between 150 and 200 questions. The size of the questionnaire is determined in part by how long it takes to administer. A good rule is that the average employee should be able to answer the questions in less than half an hour. If the questionnaire is too long, too imposing appearing, or difficult to understand, motivation will wane and some people will begin answering by making random check marks without reading the questions. To avoid these difficulties, the questions must be so phrased that only one interpretation is possible for each question. Also, the meaning of each question should be really clear. And finally, the questions usually need to be

phrased so that an adult with about six years of primary schooling can easily read and understand them. If this sounds overwhelming to you, it may make you feel better to know that even experts have great difficulty meeting all of these requirements. Propitious use of the survey committee and pretesting help overcome most of these obstacles.

Survey data can be remarkably reliable. When we consider all of the hazards so far discussed, we may begin to feel that it would take a miracle to obtain scientifically quantitative data from this instrument. Fortunately, when the above mentioned precautions are taken and the surveys are administered in the same manner, the data tends to be incredibly sound.

IT is advisable to test the survey for reliability. Reliability refers to how alike the results are from two measurements with the same survey on the same population. The more alike two measurements are with the same "yardstick" the more reliable the measuring instrument. We would really like to be reassured that the employees have answered the survey questions honestly. This means giving their honest opinions or attitudes. It does not mean giving the "right" answers to questions. Generally, there are no "right" answers. Probably the simplest way to measure reliability is to compare the employees' replies to neutral questions on succeeding surveys. Neutral questions are those for which there is no logical reason to expect change.

We shall now demonstrate two methods of measuring the reliability of surveys. For this purpose, the data from two surveys of the same company taken

two years apart will be used. In the first survey, 515 employees and supervisors participated, while 595 responded in the second survey. In each case, this amounted to 97% of the total number employed. We may therefore conclude that these samples were both representative and adequate. Since everyone except those on vacation or sick leave participates in Employee Attitude Surveys, no sampling problem arises to complicate matters.

Since the surveys were given before and after training to measure the change in the supervisors' behavior as the result of training, most questions should show some change. However, we had fourteen neutral questions which were not related to the training measures. No logical reason could be found for change occurring on these questions. The stability and reliability of these survey data were attested to by the constancy of the replies to these fourteen questions. An example will help clarify this measurement. The following neutral question appeared on both surveys:

Neutral Question:

How often would you like to know the reasons for work changes affecting your job?

1952	1954	
79%	76%	Always
12%	14%	Usually
91%	90%	Always plus Usually, insignificant change.

The change over a two year period is only one percent, a negligible amount. A one percent change could occur by chance about once in two times. Now note the change on a similar question related to training.

Training Question:

When changes are made on your job, are you told the reasons?

1952	1954	
30%	36%	Always
41%	46%	Usually
71%	82%	Always plus Usually, very significant change.

This change of 11% could occur by chance less than once in 1000 times. Nevertheless, this does not prove that the training was successful in this area. It is still necessary to prove that the trained supervisors changed more than the untrained control group by a significant amount.

We will illustrate a second method of testing reliability by another example. Below are a "change-item" and a group of items on the same subject to be used for comparison purposes.

Change-Item from 1954 Survey

During the past year, how has your supervisor changed? His supervision has become:

- much better
- little better
- about the same
- little worse
- much worse

Comparison Items 1952 & 1954 Surveys

Supervisors were rated as very good, pretty good or not so good on each of the following:

- Knows overall job
- Knows each person's job
- Gives out work fairly
- Listens to complaints and grievances

These are a sample of twenty-five items directly related to supervision which were used on both the 1952 and 1954 surveys. Each supervisor was rated by all of his employees on each item. The actual change in rating for a super-

visor was obtained by subtracting his 1952 score from his 1954 score on these 25 items. This actual rating was then compared with the rating on the single "change-item" for each supervisor. This comparison could be made for thirty first-level supervisors. An extremely high relationship was found between these two independent measures of supervision. The coefficient of correlation was 0.844. This unusually high relationship was frankly not anticipated. This very high correlation attests to a remarkable consistency in employee responses.

THE two measures of reliability discussed here gave independent proof of the unusual reliability possible in Employee Attitude Surveys. Another coefficient of reliability may be found by calculating the correlation between the responses of the work-groups to the neutral questions on two surveys. Such calculations of reliability are essential to assure you that your data are sound.

The survey gives factual information about the problems existing in the work groups. It may point to possible problems in the areas of communication, supervision, work adjustment, participation, morale, etc., depending upon the questions asked. Since one purpose of training is to improve the interpersonal relations between each supervisor and his employees, the survey data may be used to locate real problems which can serve as training material. Training of the supervisors may then take place in their natural work groups with their employees, using real problems as the curriculum.

Aside from theoretical considerations, there is the one insistent reason for util-

izing the survey data as the training curriculum. It appears that this is the only training method, so far reported, which has proved successful by scientific measures. Mahler and Monroe have made a comprehensive survey of training programs and concluded that many have acclaimed the value of training. But it appears that the supervisors have only acquired a "verbal veneer." In none of these studies has the supervisors' behavior on-the-job been shown to change. Until recently, all known quantitative studies have reported negative results for supervisory training programs. Between 1948 and 1954 two independent, long term training programs have tested the survey-feedback method of training. Both proved successful.

IF you use Employee Attitude Surveys before and after training, it will be necessary to have a number of the same questions on both surveys. These may be used to compare the behavior of the supervisors before and after training. In a successful training program, the trained supervisors will make significantly more progress on these questions than the untrained supervisors.

Thus, it is necessary to have an experimental group of trained supervisors and an approximately matched control group of untrained supervisors. It is not necessary to have large numbers of supervisors in each group when this method is used. The statistical analysis can be made to depend on the number of employees involved rather than the number of supervisors. For example, ten first-level supervisors in each group could have a total of 200 employees reporting to them. Thus, the numbers for

statistical computations would be about 100 in both the experimental and control groups.

AN easy way to handle the problem of matching supervisors for the two groups is to divide each department into an experimental and control group. In this manner, the supervisors in each group can easily be matched by supervisory level, sex, type of work, and having the same boss. Further matching by age, education, etc., is optional and should be done if convenient.

Perhaps the easiest way to measure the results of your training program is to compare the change in rating of the experimental group and the control group on each question individually. You will have between 50 and 100 questions which measure the behavior of the supervisors on-the-job as seen by their employees. These constitute an adequate measure of training.

We will illustrate this measurement with a question from our survey.

Survey Question:

My supervisor gives the worker credit for a suggestion. (check one)

	1952		1954	
Experimental group:	66%	Usually	82%	Usually
Control group:	67%	Usually	66%	Usually

From this data we note two things. First, the control group changed negligibly in two years. Second, the experimental group made progress. The total relative progress for the experimental group was 17%, because the control group went backwards one percent. Since there were 85 employees in this experimental group and 130 in the con-

trol group, calculation showed that this change could occur by chance less than once in 1000 times. Therefore, we may conclude that the experimental group made significantly more progress than the control group on this one item. By performing the same calculation for all items, we determined our final results. In the case of our training program, the experimental group made significant gains on 67% of these items while the control group was ahead on nine percent of them. On the remaining items there was no significant difference between the two groups. These results are overwhelmingly in favor of the experimental group.

Having proved that our training program worked once does not guarantee that it will work again. It is important to replicate the experiment. That is, repeat it again in almost precisely the same manner. For example, the survey-feedback training experiment has been performed independently in two public utilities with much the same results. This can be considered replication.

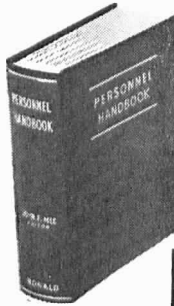
This type of experiment should also be validated by some second independent measure of supervisory behavior. It is difficult to get another direct measure, so an indirect measure will suffice. Some suggestions are: a group of higher-level supervisors may rate each trainee supervisor; measure the productivity of the groups; use absenteeism as a measure; give the trainee supervisors an attitude questionnaire.

We used the last method described. Caution must be observed here. The questionnaire must not be one which tests what the supervisors have learned

(Continued on page 48)

PERSONNEL HANDBOOK

Today's best guidance for securing, maintaining an adequate staff; avoiding labor friction; reducing turnover; making best use of employee skills. Handbook covers: organizational planning; health, medical services; morale; the personnel audit; etc. Analyzes relationships between employee and management; company and public; management and employee representatives; management, employee, and government agencies. Includes policies, techniques for every kind of business. *Edited by John F. Mee with 65 Contributing, Consulting Editors. 1167 pp.; 262 ills. \$12*



Other famous RONALD Handbooks

ACCOUNTANTS' HANDBOOK, 4th Ed.	\$15
COST ACCOUNTANTS' HANDBOOK	12
FINANCIAL HANDBOOK,	
3rd Ed., Rev. Printing	12
FORESTRY HANDBOOK	15
MARKETING HANDBOOK	12
PRODUCTION HANDBOOK	12

• Through bookstores or from:

THE RONALD PRESS COMPANY
15 East 26th St., New York 10

EVALUATE

(Continued from page 23)

and can tell about as in a "how to do it" course. This tests "verbal veneer" rather than the deeper attitudes which affect behavior. The test must tap deeper attitudes.

When two independent measures are used, we may appreciably increase the validity of our results. One well known procedure is to compare the top 25% with the bottom 25% of a group under study. As an example, we did this with the thirty first-level supervisors who

were in training. All thirty supervisors answered a Supervisory Attitude Questionnaire designed to tap deeper attitudes before and after training. The eight supervisors who made the most progress on this questionnaire were compared with the seven who made the least progress. Here are the results. First, the eight top supervisors were found to have attended almost three times as many training meetings as the seven bottom supervisors. Second, the eight were rated twice as high on the average as the seven in supervision on the Employee Attitude Survey. And finally, the employees of the eight top supervisors rated their groups 25 times as high in group-morale as the employees of the seven lowest supervisors.

Use of the survey as an evaluation instrument led to the development of a supervisory training program which succeeded in improving the supervisors' behavior on-the-job where it counts. The survey-feedback training programs which were so effective took about four years each, even though the actual successful experiments were conducted over only two years. Continued use of evaluation should make it possible to develop effective shorter term training programs. Finally, we are cognizant of the fact that some of the training programs which have failed to use scientific evaluation may have been successful. Evaluation would settle such questions. We feel that the evidence presented clearly indicates the importance of evaluating training programs.

Order your Journal Binders now. Send \$2.00 to John Skelly.