Evaluating the Training Effort

a positive approach for training directors and training program evaluations

Joseph Wolfe

Training directors have at their disposal a technique that can help them increase their effectiveness and impact within the organization. This technique, known as the experimental program evaluation, can radically alter the profile and effectiveness of the training function.

The corporate training, by utilizing this new technique, can place himself squarely in the center of organizational development--but to do so he must understand and apply rigorous evaluation techniques to his training and development programs.

Classical Experimental Ideal

After a number of pioneering studies in the 1950's, the characteristics of reliable and rigorous evaluative research design have been firmly established. In evaluating a management development program, that evaluation's ideal methodology would have all the earmarks of the classic laboratory experiment. It would have the quantified, accurate and objective measurement of change from the "before" training state to the "after," the strict identification and isolation of cause and effect, the use of statistically equivalent experimental and control groups which have been subjected to the same before and after measures, and at a minimum, a specification of the instantaneous and short-run effects of the training effort. Diagramatically, the evaluation would look like:

$$E_1 \; T \; E_2$$
$$C_1 \quad C_2$$

where the experimental group (E_1) and the control group (C_1) are statistically the same and E_2 and C_2 are those same groups at the end of the training program with only the experimental group receiving the treatment or training (T). In this logically foolproof design, any differences between E_2 and C_2 must be attributed to the training given and to no other source.[1,2]

Review of Past Studies

The evaluation procedure just noted is the ideal and theoretical absolute and serves as the model for training program evaluations. If the training director is to improve upon his efforts and performance, he must implement as rigorous an evaluation scheme as possible.

Reviewing a convenience sample of 21 of the better known and publicly available studies that have attempted a rigorous evaluation of training program effectiveness, it is apparent that they often fail to include many of the crucial elements needed to accomplish a valid and reliable evaluation.[3] The sample also illustrates what *has* been done for it is upon past efforts that today's training director makes his own progress.

Table 1 illustrates that the (1) use of controls of some type, and (2) an attempt to gauge on-the-job behavior change, has become almost standard procedure. Once past these two points, however, all sorts of methodological varieties crop up. Ideally there should be a random assignment of members to the experimental and the control group; if this cannot be accomplished, a control group should be established that has been matched to the experimental group in every relevant and distinguishable aspect.

Whatever the method for creating the control, its statistical equivalency or "sameness" to the experimental group must be determined. In our convenience sample, only two studies featured random assignment while another seven had attempted to match a control to the trained group; it could not be determined if equivalency had

Table 1.
Evaluation Study Characteristics

Study	Control	Type of Control		Measurement		Standardized and Validated Instruments	Tests of Program Inputs	Focus of Evaluation			
								Short-Term Effects		Long-Term Effects	
		Random Assignment	Matched or Equivalency	Before Training	After Training			Reaction	Learning	Behavioral Change	Operational Performance
Osterberg and Lindbom (1953)	No	-	-	No	No	No	Yes	Yes	No	Yes	No
Blocker (1955)	No	-	-	Yes	Yes	No	No	No	No	Yes	No
Fleishman, Harris and Burtt (1955)	Yes	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Barthol and Zeigler (1956)	No	-	-	Yes	Yes	Yes	No	No	Yes	No	No
Buchanan (1957)	No	-	-	No	Yes	No	Yes	Yes	No	Yes	No
Goodacre (1957)	Yes	Yes	-	Yes	Yes	No	No	No	Yes	Yes	No
Massey (1957)	Yes	-	Yes	No	Yes	No	No	No	No	No	Yes
Moon and Hariton (1958)	Yes	No	No	No	Yes	No	No	No	No	Yes	No
Sorenson (1958)	Yes	No	No	No	Yes	No	No	No	No	Yes	No
Buchanan and Brunstetter (1959)	Yes	No	No	No	Yes	No	No	No	No	Yes	No
Stroud (1959)	Yes	No	Yes	No	Yes	Yes/No[1]	No	No	Yes	Yes	No
Mahoney, Jerdee and Korman (1960)	Yes	No	No	Yes	Yes	No	Yes	No	Yes	No	No
Miles (1960)	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No
House and Tosi (1963)	Yes	No	No	Yes	Yes	Yes	No	No	No	Yes	No
Carron (1964)	Yes	No	Yes	Yes	Yes	Yes	No	No	Yes	No	No
Moffie, Calhoun and O'Brien (1964)	Yes	No	No	Yes	Yes	Yes	No	Yes	Yes	No	No
Miles (1965)	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Thorley (1969)	No	-	-	No	Yes	No	No	No	No	Yes	No
Baum, Sorensen and Place (1970)	Yes	No	Yes	Yes	Yes	No	No	No	No	Yes	No
Hayes and Williams (1971)	No	-	-	Yes	Yes	Yes	No	No	Yes	No	No
Roy and Dolke (1971)	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	No	No

[1] Test consisted of Fleishman Leadership Opinion Questionnaire and case studies

been obtained in every case.

Must Know "Before" State

In measuring the amount of change brought about, it is necessary to know the trainee's state before the training has begun—and it is best if the measurement is made with an instrument of proven accuracy. In our sample, eight of 21 of the evaluations used no before measurements, and of those evaluations that did perform a pre-training measurement, about two-thirds of them used standardized and validated test instruments. When conducting a training program, it is also useful to know which of the program's inputs or materials were the most effective. Here the concern is more with *how* the change was brought about rather than the *how much?*[4]

It might be interesting to the training director to know that after 30 hours of exposure to a comprehensive human relations program, 80 per cent of the trainee's attitude change could be attributed to an outstandingly good 20-minute BNA training film; or perhaps all the time and effort spent to create six original in-house case studies were completely wasted after it was discovered that the case method is an ineffective way of teaching management principles. In any event, only four of 21 evaluations attempted to determine which training inputs were the most productive.

Test Four Areas

It was noted earlier that most of the evaluations focused their attention on some type of behavioral change on the job. In evaluating a training program, the evaluator can and should test for (1) a reaction from at least the trainee,[5] (2) the amount of learning or attitude change that went on, (3) a difference in the way the trainee behaves once he returns from the training experience, and (4) an improvement in the working unit's operating performance.[6]

These tests also delineate the necessary parts of the successful training process. To train, the student must first be jarred or upset. After the trainee is jogged, he can be taught something. After he is taught something, the trainer can look for some outward manifestation of an internal mental change. This change in behavior or different way of acting also demonstrates that the student is practicing on the material and is less likely to lose the skill through a lack of use.

The Ultimate Test

Purely academic educators must be satisfied with achieving the first three elements. But, because this is training, the ultimate test of effectiveness must be met--namely an improvement in an indicator of operational performance. Typical operating indicators are higher profits, lower absenteeism, a higher return on assets controlled by the unit, or lower labor costs. The only reason the trainer wants a reaction is so that he can impart material, and the only reason he wants to impart material is to get the manager to behave differently.

The trainer must not be content with behavior change only, for change must be translated into something meaningful to the organization. Behavior change without operational improvement is a barren adventure. However barren this adventure, most of the evaluations surveyed looked for a change in the manager's behavior without looking for the organizationally meaningful operating improvement. The evaluations also failed to systematically[7] determine reactions or the amount of

learning that actually took place. Many studies jumped to some external indicator of reaction and learning by attaching themselves to behavior change as if behavior change proved that the trainee learned and reacted in the fashion desired by the trainer.[8]

Some Emulation Attempted

Although none of the studies completely fulfilled the requirements of the experimental ideal, in more recent years the evaluators have at least attempted to emulate the laboratory experimental model. In constructing their more rigorous evaluations, they have been aided by the development of valid paper and pencil tests, and where these have been absent, they have better theory from which they can construct their own test instruments.

From the array of evaluations surveyed, the most rigorous ones were those conducted by Miles (1960 and 1965), Goodacre (1957), Fleishman, Harris and Burtt (1955), Baum, Sorensen and Place (1970), and Roy and Dolke (1971). As a group, they featured before and after measures on both control and experimental groups, and many used standardized measuring devices. Almost all attempted to measure the trainee's degree of on-the-job behavior change. The most comprehensive evaluation was performed by Fleishman, Harris and Burtt as they measured the trainee's reaction, learning and behavior change.

Evaluations and the Trainer

The published evaluations demonstrate what has been done and how complete and incomplete the science of evaluation is at this time. It is heartening to see the continuing improvement in evaluation methodology and theory–

especially within the last 10 years.

It is unfortunate, however, that few of the evaluations were performed by those in charge of the training function. Too often the evaluations were performed by academicians and consultants on a sporadic basis rather than on a continuing basis; too often evaluation expertise was imported for the occasion.

This does not bode well for the training director for it is his present and future credibility that is at stake. The line operations of the firm have a profit and loss or budgetary system that keeps tabs on their effectiveness. Because an effectiveness system for the training department does not exist, it is up to the training director to create a viable one or be forced to accept arbitrary budgetary decisions from top management.[9]

No Justification In "Crunch"

Too often, when a budget crunch is imminent, the training director cannot justify, in benefits to the organization, monies spent on training. Under these circumstances, top management is probably justified when it cuts back in the "soft" personnel and training areas when "hard" and tangible profits can be seen emanating from plant and equipment expenditures. Faced with this situation, the harassed training director can only appeal to top-management's concern for the long-run, or "prove" training results with hastily assembled and conceptually weak *post hoc* evaluations.

The training director's situation was put into relief at a seminar recently conducted by the author. With attendees a cross-section of all industries in an urban area, one had to conclude that none of the directors were doing anything meaningful to evaluate their training efforts. While there was full

agreement over the need to evaluate the effectiveness of their training programs, there was disagreement over the focus of the evaluation and the methods to be used. They also doubted their own ability to mount a meaningful evaluation project and they doubted also anyone else's ability to do so. Ultimately they doubted that any results from the type of evaluation that *could* be performed would be worth the time or expense.

Evaluations Not Meaningful

This does not mean they were not performing any evaluations–it just meant that they were not performing meaningful ones. Most firms asked for the participant's reaction to the particular program.[10] One was using a consultant to conduct a procedural analysis of its programmed learning materials. Another firm had performed an uncontrolled substantive evaluation of two programs conducted for them by the American Management Assn.[11]

On a gradient of sophistication, all firms were using what would be termed a common-sense evaluation, a few were employing systematic evaluations in the form of reaction questionnaires, while none were using any form of experimental evaluation as outlined in this article.

A Positive Approach

The question "can evaluations be made?" is no longer the central one, for as demonstrated, some type of evaluation can be made. The question now is: "How can more meaningful evaluations be made?" In one way or another, for better or for worse, the firm's training function is being evaluated–the point now is to make evaluations systematic, objective, and productive. The experimental

ideal and its rationale have been outlined and the evaluator's progress and deficiencies have been noted. It is now up to those in control to implement and improve upon training evaluations within their organizations.

The reasons given for not attempting to evaluate are numerous: it costs too much, it is impossible to establish controls, measuring devices are not available, the staff is not qualified, the statistical work is too cumbersome and complicated, it is impossible to determine the relationship between training and results, too many variables are operating at the same time, the evaluation results are too theoretical and do not prove anything, *ad infinitum*.

These problems can be solved by an ingenious, vigorous and aggressive training director. As part of a total training project, the costs of evaluation are minor. Also, because those evaluations that *are* conducted are done on a piecemeal basis, there is little synergism or few economies of scale. Many of the other excuses can be accommodated through careful planning. As stated by Goodacre, "the design for the experimental evaluation, including criteria, controls, and statistics, should be developed as an integral part of the training program, not as an afterthought."[12]

While the training program is being formulated, the objectives of the program should be clearly stated, the most practical methods for accomplishing the objectives should be chosen, and success criteria should be selected which can be measured and are meaningful to both the operating supervisor and top management. At this time the realities of what can be accomplished with the in-house evaluation expertise will also be put into relief.

Ingenuity Needed

Rather than being humilated into apathy and defeat, the trainer can either reach outside the organization for expert advice, or thoroughly "go to school" on the literature and make applications and adjustments using other evaluations as models.[13] Whatever the course of action chosen, ingenuity and resourcefulness will have to be employed to carry the project to its end; the problems of experimental evaluation in the real world are difficult but they are not insurmountable.

For example, the author solved the problem of control in one evaluation study by publicizing the program intensively throughout the organization, thus creating an oversubscription to the program. As the course was designed to accommodate 20 participants and 41 signed up for it, a random assignment of persons into the experimental and control groups created two almost equally-sized groups with statistically identical characteristics.[14]

In the area of measuring devices, the author, rather than using a standard paper and pencil test, had to construct a focused interview schedule to determine the effects of a ghetto management training program. In this case, the task was to determine the reactions, learning and behavior changes of potential black entrepreneurs. The lack of a standard test was not a great hardship for the focused interview technique is especially appropriate in unstructured and unique situations. This evaluation was more exploratory in nature and was a necessary first step so that more rigorous program evaluations could be made in the future.

An Evaluation-Centered Strategy

Although the study and application of evaluation methodology can be pursued as a science unto itself for its own virtures, evaluations can also serve as a means to the end of organizational effectiveness. The training department can make itself a pivotal factor in achieving organizational effectiveness if it can design and implement programs that insure that the firm's manpower development programs conform to the firm's need for organizational growth and effectiveness.

As gathered from the seminar, training directors have handicapped themselves by not being able to improve their effectiveness objectively. Quite often the training director's clout within the organization has depended upon the good will of top management.

While there is no denying that many executive decisions must be made on faith alone, top management's faith can be tested only so many times. Rather than relying on faith, it is proposed that the training director rely less on exortation and more on positive, results-oriented action. What is proposed is the system as depicted in Chart 1 where evaluations play a central and action-oriented role.

Goals Identified

In this system, top or central management informs the training director of its overall organizational goals. The training director then determines the training needs. This training needs inventory must include not only those already held within the organization, but also those who are brought in from the outside.

After inventorying and determining training needs, the trainer must construct a series of training programs that hopefully will correct any deficiencies that have been discovered. When constructing the programs, the training director must build in effectiveness evaluations so he will be able to monitor and correct his own performance. After conducting a number of programs, major revisions can be made if necessary, and training results, both internal and external to the training department are reported back to top management. Within his department, the trainer can measure the short-term effects of his programs in the classroom; to measure the long-term effects, the trainer must look outside his department for on-the-job behavior change and operational results.
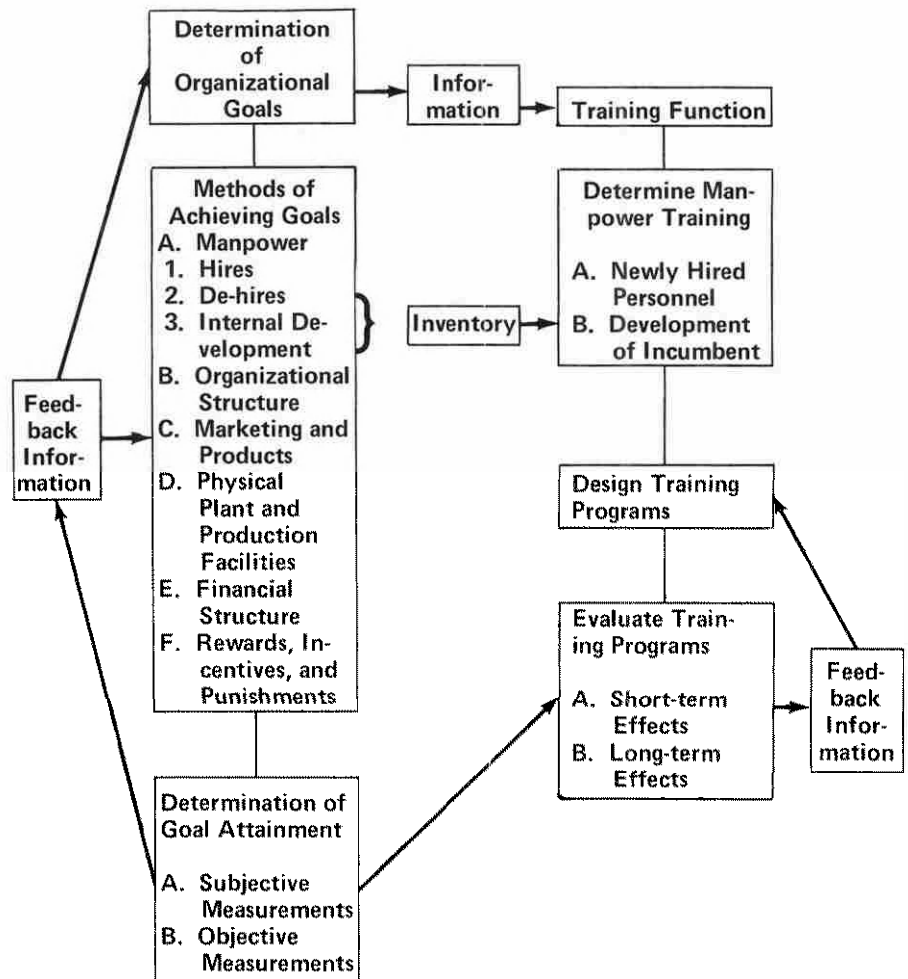
Risks Involved

This strategy, which should heighten the training department's profile, is not without risks. The

Chart 1.

An Evaluation-Centered Training System

strategy requires the training director to cast himself into an aggressive and opportunistic mold—a mold that is foreign to many trainers. The strategy is also centered on an evaluation technique that is inimical to many trainers. The strategy also requires the training director to ready himself for a number of short-run defeats and confidence shattering experiences—an anachronistic result of many training evaluations is that the more rigorous the

research design becomes, the more likely the experiment will prove that the training is failing to accomplish its purpose.

But these risks must be taken for the ultimate viability of the organization's training function. As the training department begins to acquire greater sophistication and evaluations become part and parcel of training efforts, the results of the evaluations will begin to manifest themselves throughout the organization. As

these results are felt in higher and higher places, the department's credibility will increase thus making it easier to mount even more adventuresome and meaningful program evaluations.

Conclusion

Trainers need not proceed blindly regarding the effectiveness of their training programs. The experimental program evaluation can do much to discover the degree to which training and results are actually being brought about. This does not mean that there is not much work to do until evaluations become a science. At least the theory and methodology of program evaluation is known; enough pioneering studies have been conducted to crystallize what must be done to accomplish a worthwhile evaluation. More and more of the reported studies contain controls and are using before and after measurements with standardized tests. Most of the evaluations tend not to be very comprehensive though—they often focus on only one or two aspects or effects of the training process.

If used properly, a training system, that includes consistent and comprehensively applied audits and evaluations, can do much to make the training department a more effective operating unit within the organization. The evaluations can provide the training director more solid information with which he can audit and correct his own performance. Also, if the evaluations are focused more on the long-term effects of the training, the trainer can more directly and concretely gauge and demonstrate his contributions to the firm.

In the future, the training director must conduct evaluations that will include not only an audit of classroom procedures, but also the effects of the various teaching techniques, devices and inputs, as well as the staying power of training's effects. USE RSTD

Footnotes

1. The Solomon four-group design extends the diagram to measure and control the effects of any contamination brought about by testing the control and experimental groups in their "before" state. Solomon suggests adding an E_3 and C_3 with a treatment applied to E_3. See Richard L. Solomon, "An Extension of Control Group Design," *Psychological Bulletin,* Vol. 46, No. 2 (March 1949).

2. A placebo applied to the control group would put into relief the *real* effects of the training versus the effects of the *act* of training. As Edward A. Suchman concludes though, the use of placebos in any application other than in clinical and drug evaluation studies is largely hypothetical. See the Discussion in his *Evaluative Research: Principles and Practice in Public Service and Social Action Programs,* New York, Russell Sage Foundation, 1967, pp. 96-100.

3. While not an exhaustive compilation, the studies reviewed represent some of the more ambitious evaluations performed and described in the literature over the past 20 years. Entrance into the sample was also determined by a desire to demonstrate the chronological development that has occurred in evaluative research.

4. This question (How?) can be answered within the context of the evaluation if it has been asked early enough during the formulation stage of the evaluative study.

5. Other relevant reactions would be the trainer's reaction and an outside auditor's appraisal of classroom procedure.

6. This reaction, learning, behavioral change, and operational performance schema is only a slight modification of Donald L. Kirkpatrick, "Techniques for Evaluating Training Programs," *Journal of the American Society of Training Directors,* Part I, Vol. 13, No. 11 (November 1959).

7. The word systematically is used here, for trainees are reacting whether their reactions are recorded or not. The trainer is also very much aware of student reactions from the classroom experience. What is stressed here is the systematic codification of reaction so some type of analysis can be performed.

8. Many would say that a behavior change is *prima facie* evidence that learning has occurred but actually the trainee may be acting differently because he has been singled out for training rather than actually being trained. This training effect is especially hard to separate out from the training process and has been mentioned in footnote 2.

9. In responding to a questionnaire that asked members of the American Society for Training and Development to rank the

importance of their activities, no specific reference was made to creating evaluation systems. Instead they ranked the conducting and leading of conferences of first importance and the designing and development of training programs of next importance. See Myron Roomkin, "Who Are ASTD Members?" *Training and Development Journal*, Vol. 25, No. 5 (May 1971), p. 35.

10. This conforms to the results of the Cantalanello and Kirkpatrick study which found that 77 per cent of the 154 companies surveyed measured trainees' reactions to their training programs; little else was measured. See Ralph F. Cantalanello and Donald L. Kirkpatrick, "Evaluating Training Programs--the State of the Art," *Training and Development Journal* Vol. 22, No. 5 (May 1968).

11. A procedural evaluation concerns itself with the quality of the training program's inputs while a substantive evaluation looks to the training program's results or impact.

12. Daniel M. Goodacre, "The Experimental Evaluation of Management Training: Principles and Practice," *Personnel*, Vol. 33, No. 6 (May 1957), p. 535.

13. Outside consultants should be used in an advisory capacity only, however, tempting it is to let them perform the evaluation. If consultants are used, use them for advice only, understand completely why they make the judgments they do based upon what

training and evaluation theory, and require that they leave behind them their tools so you can use 'and re-work them again in future evaluations of your own.

14. The evaluator, in this example, had to wrestle with the moral issue of depriving training to the 21 control group members who asked for training and probably needed it. In the short-run this must be done so the training all individuals in the firm receive will be better. Also once an individual has served as a control, there is no reason he cannot be offered a similar course at a later date.

References

Barthol, Richard P. and Martin Zeigler, "Evaluation of a Supervisory Training Program with How Supervise?" *Journal of Applied Psychology*, Vol. 40, No. 6 (1956).

Baum, Bernard H., Peter F. Sorensen, Jr. and William S. Place, "The Effect of Managerial Training on Organizational Control: An Experimental Study," *Organizational Behavior and Human Performance*, Vol. 5, No. 2 (March 1970).

Blocker, Clyde E. "Evaluation of a Human Relations Training Course," *Journal of the American Society of Training Directors*, Vol. 9, No. 3 (May-June 1955).

Buchanan, Paul C. "Evaluating the Results of Supervisory Training," *Personnel*, Vol. 34, No. 4 (January 1957).

____ and Philip H. Brunstetter, "A Research Approach to Management Improvement," *Journal of the American Society of Training Directors*, Vol. 13, No. 2 (February 1959).

Cantalanello, Ralph F. and Donald L. Kirkpatrick "Evaluating Training Programs--the State of the Art," *Training and Development Journal*, Vol. 22, No. 5 (May 1968).

Carron, Theodore J. "Human Relations Training and Attitudes Change: a Vector Analysis," *Personnel Psychology*, Vol. 17, No. 4 (Winter 1964).

Fleishman, Edwin A., Edwin F. Harris, and Harold Burt, *Leadership and Supervision in Industry*, Columbus, Ohio, Bureau of Educational Research, Ohio State University, 1955.

Goodacre, Daniel M. "The Experimental Evaluation of Management Training: Principles and Practice," *Personnel*, Vol. 33, No. 6 (May 1957).

Hayes, William G. and Eugene I. Williams "Supervisory Training--an Index of Change," *Training and Development Journal*, Vol. 25, No. 4 (April 1971).

House, Robert J. and Henry Tosi "An Experimental Evaluation of a Management Training Program," *Academy of Management Journal*, Vol. 6, No. 4 (December 1963).

Kirkpatrick, Donald L. "Techniques for Evaluating Training Programs," *Journal of the American Society of Training Directors*, Part I, Vol. 13, No. 11 (November 1959), Part III, Vol. 14, No. 1 (January 1960), and Part IV, Vol. 14, No. 2 (February 1960).

Mahoney, Thomas A., Thomas H. Jerdee, and Abraham Korman "An Experimental Evaluation of Management Development," *Personnel Psychology*, Vol. 13 (1960).

Massey, John C. "An Evaluation

of Postal Service Carrier Training," *Journal of the American Society of Training Directors,* Vol. 11, No. 5 (September-October 1957).

Miles, Matthew B. "Human Relations Training: Processes and Outcomes," *Journal of Counseling Psychology,* Vol. 7, No. 4 (Winter 1960).

_____ "Changes During and Following Laboratory Training: A Clinical-Experimental Study," *Journal of Applied Behavior Science,* Vol. 1, No. 3 (July 1965).

Moffie, Dannie J., Richard Calhoun, and James K. O'Brien "Evaluation of a Management Development Program," *Personnel Psychology,* Vol. 17, No. 4 (1964).

Moon, C.G. and Theodore Hariton "Evaluating an Appraisal and Feedback Training Program," *Personnel,* Vol. 35, No. 3 (November-December 1958).

Osterberg, Wesley and Theodore Lindbom "Evaluating Human Relations Training for Supervisors," *Advanced Management,* Vol. 18, No. 9 (September 1953).

Roomkin, Myron "Who Are ASTD Members?" *Training and Development Journal,* Vol. 25, No. 5 (May 1971).

Roy, S. K. and A. M. Dolke "Evaluation of a Supervisory Training Program," *Training and Development Journal,* Vol. 25, No. 12 (December 1971).

Solomon, Richard L. "An Extension of Control Group Design," *Psychological Bulletin,* Vol. 46, No. 2 (March 1949).

Stroud, Peggy V. "Evaluating a Human Relations Training Program," *Personnel,* Vol. 36, No. 6 (November-December 1959).

Suchman, Edward A. *Evaluative Research: Principles and Practice in Public Service and Social Action Programs,* New York, Russell Sage Foundation, 1967.

Thorley, S. "Evaluating and In-Company Management Training Program," *Training and Development Journal,* Vol. 23, No. 9 (September 1969).

Joseph A. Wolfe is an assistant professor of management, University of Illinois, Chicago Circle Campus. He is in the process of evaluating a series of supervisor's training programs instituted at a number of Chicago-area public hospitals. Listed in *American Men of Science,* he is a member of the American Economics Assn., the Academy of Management, the American Association of University Professors and the American Management Assn. After associations with the R.H. Macy Co., Inc. and Grand Way Stores, in New York, he went on to receive the Ph.D. from New York University, New York, N.Y. His dissertation, "Increasing Black Entrepreneurship in the Ghetto: an Exploratory Study of a Management Training Program For Harlem's Blacks," evaluates the effectiveness of the formal management training program as a device for increasing black economic development.