Evaluating Management Development Programs

John Francis Roberts
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_theses

Part of the Labor Relations Commons

Recommended Citation
Roberts, John Francis, "Evaluating Management Development Programs" (1962). Master's Theses. 1783.
https://ecommons.luc.edu/luc_theses/1783

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.
Copyright © 1962 John Francis Roberts
EVALUATING MANAGEMENT DEVELOPMENT PROGRAMS

by

John F. Roberts

A Thesis Submitted to the Faculty of the Graduate School of Loyola University in Partial Fulfillment of the Requirements for the Degree of Master of Social and Industrial Relations

February
1962
LIFE OF AUTHOR

John Francis Roberts was born in LaPorte, Indiana, September 7, 1932. He was graduated from LaCrosse High School, LaCrosse, Indiana, June 1950, and from Valparaiso University, June, 1954, with degree of Bachelor of Arts.

He attended the University of Chicago Graduate School, Social Science Division, from January 1955 to January 1957. He began his graduate studies at Loyola University in January 1958.

From May 1948 to April 1957 he held positions at various times as telegrapher with the Chesapeake and Ohio, Nickel Plate and Grand Trunk and Western Railroads. In April 1957, he was employed as claim inspector in the Personnel Department of the Illinois Central Railroad. Presently, he holds a position as labor relations staff officer with that company.
PREFACE

The author wishes to express his thanks to Mr. Hallmann, director of personnel for the Illinois Central Railroad, and Mr. Oliver, manager of personnel for the Illinois Central Railroad, for making it possible to write this thesis and his gratitude to his wife for the many hours of hard work she spent helping him.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>II.</td>
<td>THE PROBLEM AND METHODOLOGY: The Problem--Methodology--Limitations--Concluding Statement.</td>
<td>14</td>
</tr>
<tr>
<td>III.</td>
<td>PROGRAM EVALUATION SYSTEMS: CHARACTERISTICS, SCOPE AND CRITERIA: What Should Be Evaluated--The Criterion Problem--Concluding Statement.</td>
<td>19</td>
</tr>
<tr>
<td>IV.</td>
<td>TECHNIQUES FOR EVALUATING PROGRAM EFFECTIVENESS: The Use of Control Groups--Techniques for Evaluating the Attainment of Program Objectives--Actual Improvement in Job Performance--Ratings by Superiors, Peers or Subordinates--Concluding Statement.</td>
<td>27</td>
</tr>
<tr>
<td>V.</td>
<td>TECHNIQUES OF EVALUATING TEACHING METHODS: Purposes and Nature--Criterion--Evaluational Methods, General--Achievement and Other Objective Test Scores--Rating By Instructors--Rating By Trained Observers--Rating By Participant--Evaluations, Comparisons of Methods--Examples--Concluding Statement.</td>
<td>45</td>
</tr>
</tbody>
</table>
VII. GENERAL INFORMATION


VIII. PROCEDURAL EFFECTIVENESS


IX. TECHNICAL EFFECTIVENESS

Outside Programs--Inside Programs--Comparisons of Technical Effectiveness, Outside and Inside Programs--Evaluation of Teaching Methods--Technical Scores, Total--Concluding Statement.

X. FINAL SCORES AND RELATIONSHIPS BETWEEN VARIOUS FACTORS AND ACCEPTANCE AND FINANCIAL STABILITY OF PROGRAMS


XI. SUMMARY AND CONCLUSIONS


BIBLIOGRAPHY

APPENDIX I. COMPANIES IN SAMPLE

APPENDIX II. QUESTIONNAIRE
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. RAW NEGATIVE SCORES: EVALUATION OF PROGRAMS.</td>
<td>62</td>
</tr>
<tr>
<td>II. RAW NEGATIVE SCORES: USE OF EVALUATIONAL DATA.</td>
<td>64</td>
</tr>
<tr>
<td>III. EVALUATIONAL TECHNIQUES: APPLICATION AND EFFECTIVENESS.</td>
<td>67</td>
</tr>
<tr>
<td>IV. TEACHING METHODS EVALUATIONAL TECHNIQUES.</td>
<td>68</td>
</tr>
<tr>
<td>V. TYPES OF OUTSIDE PROGRAMS REPORTED BY FIFTY COMPANIES.</td>
<td>71</td>
</tr>
<tr>
<td>VI. DISTRIBUTION OF COMPANIES ACCORDING TO TYPES OF OUTSIDE PROGRAMS REPORTED.</td>
<td>72</td>
</tr>
<tr>
<td>VII. TYPES OF INSIDE PROGRAMS REPORTED BY FIFTY COMPANIES.</td>
<td>73</td>
</tr>
<tr>
<td>VIII. DISTRIBUTION OF FIFTY COMPANIES ACCORDING TO TYPES OF INSIDE PROGRAMS REPORTED BY EACH.</td>
<td>74</td>
</tr>
<tr>
<td>IX. COMBINED PROGRAMS: BY FREQUENCY OF REPORTING BY FIFTY-ONE COMPANIES.</td>
<td>75</td>
</tr>
<tr>
<td>X. PARTICIPATION: INSIDE AND OUTSIDE TYPES OF PROGRAMS, BY FREQUENCY.</td>
<td>76</td>
</tr>
<tr>
<td>XI. DISTRIBUTION OF COMPANIES BY CORRESPONDING TYPES OF INSIDE AND OUTSIDE PROGRAMS.</td>
<td>77</td>
</tr>
<tr>
<td>XII. RELATIVE EXTENT TO WHICH INSIDE AND OUTSIDE PROGRAMS ARE EVALUATED BY FORTY-SEVEN COMPANIES.</td>
<td>78</td>
</tr>
<tr>
<td>XIII. EVALUATIONS OF TYPES OF INSIDE AND OUTSIDE PROGRAMS: DIFFERENCES IN FREQUENCIES.</td>
<td>79</td>
</tr>
<tr>
<td>XIV. DISTRIBUTION OF THIRTY-SEVEN COMPANIES BY EMPLOYMENT OF PERSONS WITH ADVANCED DEGREES.</td>
<td>81</td>
</tr>
<tr>
<td>XV. DISTRIBUTION OF COMPANIES BY PREFERENCE IN ACADEMIC DEGREES.</td>
<td>82</td>
</tr>
</tbody>
</table>
XXXII. INSIDE PROGRAM EVALUATIONAL TECHNIQUES REPORTED BY FORTY-ONE COMPANIES... 107

XXXIII. DISTRIBUTION BY CRITERION, OF TECHNICAL SCORES FOR EVALUATIONAL METHODS OF OUTSIDE PROGRAMS REPORTED BY FORTY-ONE COMPANIES... 108

XXXIV. DISTRIBUTION OF TECHNICAL SCORES FOR INSIDE PROGRAM EVALUATIONAL METHODS REPORTED BY FORTY-ONE COMPANIES... 109

XXXV. PERCENTAGE OF COMPANIES USING SPECIFIED EVALUATIONAL TECHNIQUES IN CONJUNCTION WITH OUTSIDE AND INSIDE PROGRAMS... 111

XXXVI. EVALUATIONAL TECHNIQUES REPORTED BY THIRTY-FIVE COMPANIES WHO EVALUATE BOTH INSIDE AND OUTSIDE PROGRAMS... 113

XXXVII. AVERAGE TECHNICAL SCORES, BY CRITERION, FOR TECHNIQUES USED TO EVALUATE INSIDE AND OUTSIDE PROGRAMS... 114

XXXVIII. DISTRIBUTION OF TECHNICAL SCORES FOR OUTSIDE AND INSIDE PROGRAM EVALUATIONAL METHODS... 115

XXXIX. TYPES OF TEACHING METHOD EVALUATIONAL TECHNIQUES REPORTED BY THIRTY-TWO COMPANIES... 116

XL. DISTRIBUTION OF SCORES FOR TEACHING METHOD EVALUATIONAL TECHNIQUES OF THIRTY-TWO COMPANIES... 117

XLI. DISTRIBUTION OF CONVERTED TECHNICAL SCORES OF FORTY-ONE COMPANIES... 118

XLII. AVERAGE CONVERTED TECHNICAL SCORES FOR FORTY COMPANIES DISTRIBUTED ACCORDING TO PROCEDURAL SCORES... 122

XLIII. DISTRIBUTION OF COMBINED SCORES FOR FORTY COMPANIES... 123

XLIV. DISTRIBUTION OF SIXTEEN COMPANIES BY PERCENTAGE OF ADVANCED DEGREES AND PROCEDURAL, CONVERTED-TECHNICAL, AND COMBINED SCORES... 125

XLV. COMPARISON OF AVERAGE SCORES OF SIXTEEN COMPANIES HAVING PERSONS WITH ADVANCED DEGREES WITH THOSE OF NINETEEN COMPANIES HAVING NO PERSONS WITH ADVANCED DEGREES... 126

XLVI. ACCEPTANCE SCORES OF THIRTY-NINE COMPANIES AND FINANCIAL STABILITY SCORES OF THIRTY-SEVEN COMPANIES BY PROCEDURAL SCORE STANDINGS... 127
XLVII. ACCEPTANCE SCORES OF THIRTY-NINE COMPANIES AND FINANCIAL STABILITY SCORES OF THIRTY-SEVEN COMPANIES BY CONVERTED-TECHNICAL SCORE STANDINGS .......................................................... 128

XLVIII. ACCEPTANCE SCORES OF THIRTY-NINE COMPANIES AND FINANCIAL STABILITY SCORES OF THIRTY-SEVEN COMPANIES BY COMBINED SCORE STANDINGS ........................................................................... 130

IL. THIRTEEN COMPANIES LISTING AS FIRST CHOICE WORK EXPERIENCE IN PROFESSION-RELATED AREAS COMPARED WITH FIFTEEN COMPANIES LISTING AS FIRST CHOICE WORK EXPERIENCE IN BUSINESS-RELATED AREAS BY ACCEPTANCE, STABILITY, AND COMBINED SCORES AND PERCENTAGE OF ADVANCED DEGREES ........................................................................ 132

L. THIRTEEN COMPANIES LISTING AS FIRST CHOICE DEGREES IN PROFESSION-RELATED AREAS COMPARED WITH EIGHT COMPANIES LISTING AS FIRST CHOICE DEGREES IN BUSINESS-RELATED AREAS BY ACCEPTANCE, STABILITY, AND COMBINED SCORES AND PERCENTAGE OF ADVANCED DEGREES ......................................................... 134
CHAPTER I

INTRODUCTION

The Role of Management Development. Using an analogy which does not stretch the imagination unduly, some writers have compared the modern corporation with a living organism. The corporate organism may be characterized by two primary activities—one, because it is directed toward expansion, may be termed aggressive and the other, because it is directed toward self-renewal and self-perpetuation, may be termed protective.

Protective, and to a lesser extent, aggressive activities require the presence of a supply of managerial personnel to fill higher managerial positions as they are created or as vacancies occur. When a corporation expands it needs more managers. Even when it is relatively stable, vacancies occur from death, retirement, or resignation. If the corporate entity is to survive, it must renew itself by replacing lost managers. Thus, in order to carry out its primary activities, the corporation must, in some way, have access to a source of managerial personnel located either outside itself or, as is increasingly the case, within the corporation itself.

The effectiveness of the corporation in both its aggressive and protective activities depends upon the quality of its present management. The expanding company expands partly because the men running it know how to keep it expanding. The stable company survives because the men in control know how to protect it against competition. Thus, not only is it important for the corporation to have qualified men ready to step up to newly created posi-
tions or fill old ones, but it is equally important for it to take steps to see that the performance of those presently holding managerial positions improves as much as possible.

All of this adds up to one thing. The corporation must be in one sense or another an educator. It must train men to perform their work more effectively; it must prepare them to accept greater responsibilities; in short, it must, as a matter of logical necessity, teach and educate. If it does not, it will soon perish.

The corporation has functioned in its role of teacher and educator with varying degrees of success. In days past, it did so, haphazardly--men were frequently thrown into positions of greater responsibility and left there to sink or swim. Often this constituted the sole content of the corporation's role as educator. It rarely bothered to give as much serious thought to improving the performance of the management as it did, for example, to improving the techniques of production.

However, relying entirely upon work experience as a training technique has become pretty much a thing of the past. After the corporation rationalized its productive processes, it turned its attention to management and began to rationalize the training of management. The result has been that an ever increasing number of corporations have developed formal programs specifically aimed at training management people and improving their effectiveness.

The growth of formal management development programs has come about rapidly. According to one observer, the number of "in-plant" management development programs in 1943 were so few that one large national association
did not feel it worth its while to make a survey of them. By 1956, the Mutual Benefit Life Insurance Company found that 70.2% of all American corporations conducted formal management development programs and that another 16.3% were considering initiating programs. In 1960, it was estimated that American corporations spent more than twenty million dollars on "outside" management development programs alone.

Taken by itself, this might be interpreted to mean that corporations have accepted management development programs without qualification. There are indications that this may have been true at one time. Daniel Goodacre, for example, wrote that although management expects a return from training, it does not make anything approaching the kind of effort it makes to calculate precisely the returns from manufacturing and sales. Raymond L. Randall describes corporate acceptance of management development programs as being so uncritical that they, "have tended to become a fad." According to him, "many companies have been sending executives to them, simply because it is the thing to do."

Criticism of Management Development Programs. Today this blind type of acceptance is disappearing rapidly if, indeed, it ever existed. According

---

2"Companies Sound Off on Executive Training," Dun's Review and Modern Industry, LXVII (June 1956), 90.
5Randall, p. 43.
to a survey made by Mutual Benefit Life Insurance Company, less than half (46.1%) of the companies who reported having management development programs said they were satisfied with them; 15.7% indicated qualified satisfaction; while 38.2% showed marked dissatisfaction. 6

The new, harsh attitude towards management development programs uncovered by the Mutual Benefit survey has found its way into print. John F. Chapman, for example, quotes the president of a leading corporation as saying, "We continue our program as a hedging process--for insurance purposes only. We have no evidence that it really develops executives." 7 Louis A. Allen goes further. He states flatly that there are few, if any accomplishments to show for all the effort spent on management development programs. 8 Erwin K. Taylor aims his criticism straight at training directors who, according to him, often gets by entirely on bluff and personality. He describes them thus:

One characteristic that seems to be common to the members of this group is that they are almost all highly articulate. In the main, they possess a great deal of poise and are skilled in making a presentation to top management and in addressing a small group. Many (perhaps to bolster their own sense of inadequacy) write 'little Jack Horner' articles for management magazines and trade publications. 9

Even such a sympathetic observer as Melvin Anshen feels that, "we have

6"Companies Sound Off on Executive Training," p. 90.
8Louis A. Allen, "Does Management Development Develop Managers?" Personnel, XXXIV (September 1957), 18.
too little understanding of what actually goes on in executive development programs, too little knowledge of what these programs are really accomplishing." He summarizes by saying that the pressing need of today is to learn whether, "present courses adequately or efficiently meet the need" and if "the best use is being made of the money, time, and talent invested."¹⁰

Argument Against Possibility of Evaluation. Although there exists a consensus about the desirability of systematically evaluating formal management development programs, there is disagreement on a more fundamental question, viz., is it possible to evaluate formal management development programs in such a way as to achieve meaningful data? At least one writer, Kenneth Andrews, argues most persuasively that it is not.

Andrews recognizes the importance of determining the deficiencies of management and, where appropriate, remedying them by formal programs. However, he states that the variety of needs of individual trainees is so great and that the number of factors, both known and unknown, at work to advance or retard general development is so large that it is impossible to (1) determine when the needs of a given trainee have been met and (2) isolate the results produced by formal programs from those which come from other areas of management development.¹¹

These conclusions seem to flow from a broad (though unwritten) definition of education used by Andrews. Education he sees fundamentally as the develop-


ment of the individual in all human aspects. The educated man, as thus defined, does not necessarily learn individual skills useful in coping with specific situations; rather he acquires a frame of reference which enables him to know what to look for to cope with any type of problem.

Thus the proper objective of the educational process becomes an extremely general one. When it is attained, it is the outcome of not just one or a few formal programs. Rather, it is the culmination of the individual's entire formal and informal educational experiences. Inasmuch as the objective of the educational process has not been achieved until the person undergoing the process achieves this general adaptive frame of reference, it is impossible to say whether any particular program, given prior to that time, has contributed anything to his development. In fact, unless the objective is achieved, the most that can be said for any given program is that it may not have been responsible for the failure. If the objective is achieved, the most that can be said for any program is that it may have contributed to the final success. However, it remains impossible to go back and separate the effect of a single program from the multitude of factors which contributed in varying degrees to the final product of the educational process.

If Andrews' assumptions about the nature of education are accepted, his conclusions naturally follow. Until the educated man emerges out of the student, it is impossible to tell whether any given program has been of use. After the objective of education has been achieved the number of factors involved in reaching it is so great that it is impossible to say which contributed what. It is only possible to say that all factors, considered together, affected the end result.
Argument For Evaluation. It may be well to argue that the ultimate end of education ought to be the development of philosopher-kings--men who have learned to learn and to pursue and find the truth in any situation. This might serve as the ultimate criterion of success for any extended program of education. The fact remains, however, that intermediate criteria may be used. Any single course of instruction deals with a definite, limited subject matter. If a definite, limited, concrete, subject may be taught (out of deference to the teaching profession it is assumed that it can be taught) then those being taught should know more about the subject after being taught than they did before being taught. Since what they were taught hardly approaches the infinite, it should be possible to examine them and determine how much they learned of what they were taught. Any instructor, for example, knows that his students have learned something when they give him information in an examination which they did not possess prior to beginning the course.

Since the subject matter of any given course of instruction is definite, limited, concrete and otherwise finite, it will (or should) be applied to definite, limited, concrete and otherwise finite situations. Its application to such situations should produce tangible results. Examination of tangible results should tell the observer something about how well the program succeeded in getting across its point.

One or two examples might serve to clarify this point. Characteristic-

12See R. L. Thorndike, Personnel Selection: Test and Measurement Techniques (New York, N. Y., 1949), p. 358, where the author points out that although it is almost impossible to obtain ultimate criteria, immediate and intermediate criteria will yield quite satisfactory results.
ally, human relations programs have as their ultimate objective the improve-
ment of relations between superiors and subordinates. There exist certain
well-known indices to the quality of superior-subordinate relationships.
Among these are quit-rates, absentee rates, grievance rates, suggestion rates
(where there are suggestion systems), and transfer rates. A training director,
who after conducting a class in human relations, finds that the subordinates
of the trained group quit, transfer, or are absent less frequently and, in
addition, file fewer grievances and make more suggestions than they did before
their supervisors were trained, can say without too much fear of being wrong
that the program was specifically beneficial on these points. He can speak
with even greater assurance if an untrained group of supervisors did not
show a similar improvement over the same period of time. If, on the other
hand, he finds improvement in some areas and none in others, he will know
precisely what parts of his next program should receive greater emphasis.

It is considerably more simple to learn the results of less generalized
programs. Take, for example, a program designed to reduce the number of
union contract violations by supervisors. The degree to which the program
has succeeded may easily be determined by decreases in the ratio of valid
grievances to all grievances filed.

This does not pretend to come anywhere close to exhausting the number of
possible examples. However, the two just cited should suffice to show the
feasibility of evaluating the results of formal management development pro-
grams. All that must be kept in mind is to stay away from general ultimate
criteria and ask no more of a formal program than that it achieve the speci-
fic and rather modest objectives it was designed to achieve.
Benefits of Evaluation. Assuming, on the basis of the preceding discussion, that some sort of objective program evaluation is possible, one might ask what are the benefits of carefully planned evaluations. The answer would be much along these lines.

Evaluation is a matter of logical necessity; that is, it takes place even if explicit plans have not been made for it. Participants in the program will form opinions about what it did for them; the instructor, will also have some sort of idea about the effectiveness of the program; so will the training director; and, most important for the ultimate fate of the program so will someone with the power to have it continued or discontinued. If the latter's evaluation is based on his own impressions and hearsay, to quote Goodacre, "the security of training programs, training directors and training dollars is subject to personal whims and undocumented opinions."¹³

Moreover, if only impressionistic or hearsay evaluations are made of training techniques, the quality of instruction is as likely to deteriorate as improve. In the absence of experimental evaluations, the training director has little to go on except intuition in deciding which methods and training aids are best for accomplishing program objectives. A study conducted by Walter R. Mahler points up this fact. According to him, in the absence of experimental evaluations, the teaching methods and training aids used by an organization tend to be those:

1. The training director likes;

2. Which have been used somewhere else and reported to be successful; or,

¹³ Goodacre, p. 535.
3. Which trainees tend to like.  

Conversely, if the evaluation is based on objective evidence collected in accordance with a carefully designed experimental plan, the strengths and weaknesses of the program will become readily apparent. Weak points then may be strengthened; irrelevant ones eliminated and, on occasion, a new program may be substituted for one which has proved faulty. In short, those charged with responsibility for the program will know "what is needed; what is happening; what is likely to happen; what actually happened; what still needs to be done, and be able to act accordingly."  

One might reasonably expect that the use of experimental evaluations should result in more effective use of available funds. This expectation has borne out by a number of reports appearing in management development and training literature. In one instance, an experimental evaluation made and reported by Fryer and Edgerton of six and eight week training courses for Air Force personnel showed that the eight week training course could be dropped inasmuch as students learned and retained no more information from it than from the shorter course. In another instance, Baxter, Taffee and Hughes, after making a carefully planned evaluation of three types of

---


17 It is well to note that as a general rule, evaluations conducted by insurance companies stand head and shoulders above those conducted by other organizations. This reflects, perhaps, the greater experience which insurance personnel have in the use of statistical methods.
training programs for debit insurance agents, found that there were no significant differences in the results produced by the different programs. They concluded that the more expensive "away-from-the job" programs could be abandoned. Still another such case is reported by Buchanan. During the course of evaluating a training program, he found no relationship between trainee ratings of that part of the program dealing with company policies and procedures and subsequent trainee performance as evaluated by supervisors. He said that consideration was being given to eliminating that portion of the program.

The major result of experimental evaluation, however, is more than just effective use of funds. It is as Anshan succinctly puts it, that it becomes possible for one to tell whether the training program is "on target." In this particular evaluation three training methods used by an insurance company were compared. The first consisted of a ten day training conference at a central school before the trainees went on the job, followed by text and problem assignments for the remainder of the year. The second involved one week of individual training by the district sales manager before the trainee went on the job, six months following text and problem assignments, one week at a training conference conducted at the central school, with the balance of the year spent on the job following text and problem assignments. The third involved one week of individual training by the local district agent prior to the time the trainees went on the job with the balance of the year spent on the job following text and problem assignments.

Two matched samples composed of trainees were drawn from each district and trainees were assigned randomly to the different programs. After completion of program, each trainee group was evaluated on (1) knowledge, (2) job satisfaction, (3) production, (4) quit-rate and (5) supervisors' ratings.


When this is not known, a training program may be praised or blamed for things for which it is not responsible or, even worse, it may fail completely to accomplish its objectives without anyone being the wiser. The assumption that one can expect any training program, no matter how poorly taught, to produce at least minimal favorable changes in the participants simply is unsafe to make. In a study conducted by Mahler, no significant changes were found in the participants of human relation classes. In other instances, programs have been shown to produce results contrary to those they were supposed to produce. William and Arnold Form found that instead of "indoctrinating the trainees with pro-management attitudes," a foreman training program "engendered hostility in them, and stimulated even greater suspicion among the workers in the shop." Fleishman found that a human relations training program for foremen actually resulted in foremen showing less consideration toward workers. In brief, it is


White reports cases where evaluational studies showed that training programs were receiving praise or blame for matters they had not accomplished or for which they were not responsible. In one instance, high sales resulted in a sales training program receiving very favorable attention. When the matter was studied further, it was found that salesmen, for the most part, were not using techniques taught them in the training program. In another instance, a training program was blamed for failure to significantly lower the turn-over rate among sales personnel. After an evaluation was made, the blame was placed where it belonged—on poor supervision and bad company policy.

22Mahler, pp. 294-295.


never correct to conclude, a priori, that any training program is beneficial to some extent. It must be proven, not merely assumed, to be "on target."

**Concluding Statement.** It has been shown that formal management development has become an increasingly vital part of today's corporate activity. However, in the process of doing so, it has been subjected to criticism on the grounds that insufficient efforts are made to evaluate and improve its effectiveness.

The benefits to be derived from successful evaluations are many—more effective training, more economical uses of training resources, and increased confidence in training activities, to mention a few. However, one writer, using an ultimate criterion of development as the standard for measuring program success, casts doubt upon the possibility of making evaluations which will yield meaningful data. In response to this, it has been shown that although evaluations may shed little light upon the development of any given individual into something approaching the classical "educated man," they will produce useful data when they are made in terms of intermediate criteria established as objectives for any given formal program.

In summary, it must be conceded that experimental evaluations are feasible when made in terms of specific program objectives and that important benefits may be derived from evaluations made of specific formal management development programs.

Having established the relevancy (and perhaps the usefulness) of the study subject of this paper, we shall now pass to the technical problems posed by the study.
CHAPTER II

THE PROBLEM AND METHODOLOGY

The Problem. The problem subject of this paper is to determine the extent to which the one hundred largest (from the standpoint of sales) U.S. manufacturing corporations evaluate the effectiveness of formal management development programs\(^1\) conducted or participated in by them and the effectiveness of techniques used by them to evaluate programs.

A secondary problem is to determine (1) the relationship between the degree of professionalization of management development training (the number of persons in a given management development department who hold advance degrees in work-related fields) and the use of evaluational techniques; (2) the relationship between the degree of professionalization and formal program acceptance and stability; and (3) the relationship between the use of evaluational techniques and formal program acceptance and stability.

In connection with the secondary problem these three null hypothesis may be advanced:

1. There is no correlation between the degree of professionalization in a given management development department and the effective use of

\(^1\)The term management development program was defined for the purposes of this study, as;

"formal program, akin to class work, designed to improve managerial performance at present level or to prepare individuals for greater managerial responsibilities, . . . other than one leading to a formal degree." (See Appendix II.)
evaluational techniques.

2. There is no correlation between the degree of professionalization in a given management development department and (a) the acceptance of formal programs by top management and (b) the financial stability of formal programs.

3. There is no correlation between the effective use of evaluational techniques and (a) the acceptance of formal programs by top management and (b) the financial stability of formal programs.

In the course of this investigation, it will be determined whether the foregoing hypotheses are correct.

Methodology. Using the July 1961 Fortune magazine ratings, a list was prepared containing the names of the one hundred U.S. manufacturing corporations which ranked highest in terms of sales. "APPENDIX I" is a copy of the list.

A questionnaire and a sheet of instructions were drafted. The questionnaire was divided into the following sections:

1. General information about recruitment policies of management development departments and academic achievement and specialization of department members.

2. Evaluation of effectiveness of "outside" management development programs.

3. Evaluation of effectiveness of (a) "inside" management development programs and (b) teaching methods.

4. Acceptance and financial stability of management development programs.
The sheet of instructions contained definitions of terms used in the questionnaire, prescribed procedures to be used in answering questions and specified that the person answering the questionnaire either be in charge of management development or the immediate subordinate of the person in charge of management development.

The questionnaire was presented to the author's advisor and others for criticism. Certain clarifying changes were made. Ambiguities which could not be resolved by rephrasing the questionnaire were clarified by definitions made in the sheet of instructions.

After the questionnaire and sheet of instructions were finalized, a cover letter was drafted and submitted for approval to Mr. E. H. Hallmann, Director of Personnel for the Illinois Central Railroad. The questionnaire, accompanied by the approved cover letter written on Illinois Central letterhead paper over the signature of the author, was mailed to the companies selected for study. Three weeks after the questionnaire had been mailed, a tracer was sent to companies which had not yet responded. A copy of the cover letter, the sheet of instructions and the questionnaire are found in "APPENDIX II."

A scale was developed for the purpose of rating the effectiveness of program evaluational systems used by respondent companies. The scale rates each system separately on procedural completeness and technical efficiency. It is not possible to say more here about the rating scale because the number of complicating factors involved in constructing it is so large that it is necessary to make them and the resultant scale the subject of separate chapters.
Limitations. The present sample, as samples go, is relatively small. Moreover, it is not representative because of the type of companies involved. However, since the companies studied are among the largest and most prosperous in this country, it can probably be said, with some degree of accuracy, that the average state of evaluation for industry, in general, is no higher and almost certainly lower than that found to exist in the sample.

Using questionnaires as a method to collect information presents a number of difficulties. While there is control over selecting the sample, there is no control over who will and will not respond. It is not unlikely that companies who failed to respond, did so for reasons which, if known, might affect conclusions made by the study. Moreover, it is difficult to draft a questionnaire in such a way as to make it understood uniformly. This too can be a source of error; however, unless serious mistakes were made in drafting the questionnaire, it should not contribute to systematic bias.

Much of the information obtained by the questionnaire is only approximate. Far better results could have been obtained had it been possible to ask detailed questions about frequencies of evaluations, amounts spent, and so forth. Unfortunately, had this been done, the questionnaire would have grown to such proportions that most companies would not have bothered to answer it. Consequently, the information produced by the questionnaire represents a compromise between what is desirable and what it is possible to obtain. Answers to the questionnaire will not show how frequently each separate evaluational technique is used; they will merely show which, of a number of techniques, are used. They will not show what proportion of the training budget is ear-marked for each type of program; they will only show which types of programs are included in the training budget. They will
not show the exact proportion to which the training budget is cut relative to other company expenditures; they will merely show in a qualitative way the approximate proportion of the reduction.

Concluding Statement. The problem subject of the paper is to determine: The extent to which the one hundred largest U.S. manufacturing corporations evaluate formal management development programs; the effectiveness of evalua-tional systems used by them; and the interrelationships between the degree of professionalization, the use of evaluations and the degree of acceptance and stability of management development programs.

The sample covered by the study is neither large nor representative. Accordingly, conclusions drawn by the study, with one exception, (that the level of evaluation found in the rest of industry is no higher than that found in the sample) cannot be said to hold true for all of manufacturing industry.

Returns to a questionnaire designed by the author and sent to companies in the sample accompanied by a cover letter written on stationary bearing the letterhead of the Illinois Central Railroad, make up the primary source of information for the study. The questionnaire method poses difficulties—there is no way in which to insure response from each company in the sample and it is difficult to formulate questions free from all ambiguity. Moreover, in order to keep the size of the questionnaire within manageable limits, it was necessary in many instances to use questions which yield approximate rather than exact information.

The relative effectiveness of the procedural and technical aspects of eval- uational systems were rated in accordance with a scale developed by the author. The factors involved in determining the effectiveness of evaluational systems and in devising a rating scale will be discussed in the next four chapters.
CHAPTER III

PROGRAM EVALUATION SYSTEMS:
CHARACTERISTICS, SCOPE AND CRITERIA

What Should Be Evaluated? Peter Castle writing in Occupational Psychology, says that two questions can be asked about any formal management development program—"First, What is its value?" and "Secondly, . . . what teaching methods are the most effective?" It is to these questions that evaluations must supply answers.

The best teaching methods known cannot improve managerial performance when they are used to present trivial or irrelevant material. Conversely, a program filled with important material loses much of its value if it is taught poorly. The goal of any program must be to present important material in the most effective manner. Program evaluation, therefore, has for its purpose determining (1) whether the program, itself, is of value and (2) whether the most effective methods of presentation are being used in the program. Evaluations which consider only one aspect of the program while neglecting the other must necessarily be considered incomplete.

The program evaluation must be technically correct. A criterion or criteria for the program must be established; proper instruments must be selected to measure progress toward the criteria; and controls must be used to hold constant as many variables as possible.2

---


2 Goodacre, p. 534-538.
The program evaluation system should have breadth as well as depth. It is not enough that comprehensive evaluations are made sporadically. All types of formal programs should be evaluated more or less regularly in order to maximize the benefits to be derived from evaluations.

Finally, conclusions drawn by a program evaluation should not be left to gather dust. They must be put to work. Where weaknesses are shown to exist, they should be corrected. Programs or part of programs which evaluations indicate are of little value should be discontinued or changed. When one training method proves superior to another or equal to a more expensive one, the superior or less expensive method should supplant the other.

To conclude, a satisfactory program evaluation system is characterized by the following administrative procedures:

1. Evaluation of all types of programs conducted or participated in by the company;
2. Evaluation of teaching methods; and the
3. Use of evaluation data as the basis for making changes in programs or teaching methods.

Over and above administrative procedures, the evaluation system must be technically correct—that is, proper instruments, with controls, should be used within the framework established by sound administrative procedure to measure progress toward well-defined criteria.

Only after both the procedural and technical aspects have been taken into account, can any evaluation plan be termed complete. Conversely, any evaluation plan which falls short of this, is deficient to some degree and must be considered incomplete.
Using this as the background, we will proceed to the problems involved in determining what goes into making up a technically correct evaluation system.

The Criterion Problem: Before one can measure, there must be something to measure. Before improvement takes place, that which is to be improved must first be determined. Anshan, speaking about evaluation, puts it thusly: "To evaluate development programs, management must have the objectives it is seeking clearly imminent."3 The act of evaluating a management development program constitutes measurement of the progress of the participants toward a given objective. Unless objectives have been established, the act of evaluation cannot take place. Thus in order to evaluate a management development program, the objectives of the program must be clearly defined in advance.

By and large, the selection of criteria for a program is a matter of judgment.4 Someone must decide what the program should accomplish. The success of the program will then be judged by the extent to which program participants approach the criteria.

It is evident that at least part of the criteria selected should be related to improvements made by the program in company operations. But what if they aren't? Should this be the case, the program, while adjudged a success, might in reality be a failure. One of the main problems, then, is to find and use criteria which will indicate the degree to which a given program affects company operation.


Enell and Haas divide evaluational criteria into three general groups: quantitative standards, where performance is judged by grievances, absenteeism, rejections, accidents, and turnover; qualitative standards, where the employee is judged by the degree to which he possesses qualities thought to be necessary or desirable for persons holding managerial positions; and verifiable standards where an employee is judged on whether he has satisfactorily performed certain acts he is supposed to perform.5

All three types of criteria are used to evaluate the performance of managerial personnel and all may be used, with varying degrees of success, to determine the effectiveness of any given management development program.

Qualitative standards may be termed "attitude-oriented." That is, they are grounded on the assumption that certain attitudes or personal characteristics are essential to success in managerial work. Where qualitative standards are used to evaluate management development programs, the programs usually have as their objective the improvement of those attitudes or characteristics considered relevant to managerial success. Thereupon, the effectiveness of management development programs established to improve managerial qualities is measured in terms of the extent to which participants improve in "initiative, cooperativeness, creativity, teamwork," etc., etc.

There are three major objections to using qualitative criteria as the basis for evaluating management development programs. The first is that they are vague and liable to be interpreted differently by different observers. As Enell and Haas put it, "They do not refer to statistical quantities and,

---

in fact, they really provide no way in which the individual's superior could readily ascertain whether the standard has been reached. If there is no way in which an observer can be certain that the standard has been reached, then one can rest assured that the confusion will be multiplied when, as usually is the case in program evaluation, a number of observers are used.

The second objection is that no one has really tried to establish firm correlations between qualitative standards and actual job performance. Until this is done, qualitative standards amount to little more than a guess about what characteristics are needed to handle a given job.

The third objection is that qualitative standards relate primarily to attitudes or personal characteristics rather than actual job performance. While it might be argued that possession of a given attitude or characteristic is a necessary condition for performance of a given act, it does not follow that the act will actually be performed. An attitude or characteristic is a mere potentiality. It reflects, at most, what might happen, not what actually does happen.

Qualitative standards are under vigorous attack from some quarters and there are indications that they are being abandoned in favor of quantitative and verifiable standards—criteria which are based directly on actual job performance. The argument in support of such criteria is summarized quite effectively by Knell and Haas' statement that instead of judging a manager, "by his inherent personality, his potential ability to get along with other people, the experience he could bring to bear, and other characteristics

\[6\text{Ibid., 24.}\]
which he may conceivable fail to use..." he should be "judged by what happens as a result of his work."\footnote{Ibid.}

Needless to say, actual job performance is the most logical criterion to use in evaluating the performance of individuals. After all, the well-being of the corporation depends upon work being adequately performed. What, then, could be more sensible than to evaluate managerial personnel according to how well they perform their work? If actual job performance is the best measure of the worth of a manager, then improvement in actual job performance by participants of management development programs is the best measure of the value of programs.

The effectiveness of individual training programs, like the effectiveness of individual managers, is best judged by what happens on the job. Thorndike for example, says flatly that objective performance is the ideal criterion for all types of training research. He points out that using job performance as a criterion avoids the problems created by qualitative criteria and "minimizes the possibility of observer unreliability or observer bias entering in to attenuate or prejudice the conclusions."\footnote{Robert L. Thorndike, Personnel \textit{Selection: Test and Measurement Techniques} (New York, 1949), p. 137.}

Individual training programs have specific, limited objectives. They are designed to improve work performance in certain prescribed areas. One of the criteria by which the effectiveness of a given program should be measured, therefore, is work performance in the area with which it deals.

Is improvement in work performance the only relevant criterion?
Probably not. There appear to be at least two additional criteria which should be used. Improvement in work performance will show that certain information imparted by the program has been placed in use; however, it will not show, for example, what or how much information was transmitted to the participants. It is essential to know what the participants have learned. When this is not done there exists no sound basis for determining what parts of the program could stand better presentation or greater emphasis. Thus a second criterion for evaluating a management training program is the amount of information transmitted to the participants.

An essential condition for the success of any training program is that it be acceptable to the participants. Its relevance must be demonstrated to them; it must be presented in an interesting fashion; it must catch their attention and motivate them to learn. However, although improvement in job performance and the level of transmission of information will show that the program is functioning adequately, they will not show whether it is possible to improve the program by raising its acceptability. Therefore, separate and apart from improvement in job performance and the transmission of information, the program must be evaluated for acceptability. Acceptability, then, constitutes the third criterion by which the effectiveness of training programs should be evaluated. 9

Concluding Statement. A program evaluational system, in order to be procedurally complete must (1) evaluate all programs with which the company is involved, (2) evaluate teaching methods where they are within the control

of the company, and (3) use evaluational data as the basis for decisions about modification or abandonment of programs, parts of programs, and teaching methods.

The methods used to make evaluations must be technically correct: That is, criteria for the program must be established; proper instruments must be selected to measure progress towards the criteria; and controls must be used to hold constant as many variables as possible.

In connection with the establishment of program criteria, it has been shown that the value of a training program is best measured in terms of the improvement it causes in job performance. However, other criteria are needed. A training program creates improvements in job performance by imparting relevant information to its participants. It can be improved only if it is known what areas of information are not covered properly. Therefore, after the overall effectiveness of the program has been determined by using the "improvement in job performance" criterion, it must be further determined whether the information furnished by the program is complete and adequately presented. This can be done only by measuring the amount of information transmitted to the participants. Finally, before any program, no matter how relevant its material, can be deemed a success, it must be acceptable to the participants. Consequently, it is important to determine the acceptability of the program.
CHAPTER IV

TECHNIQUES FOR EVALUATING PROGRAM EFFECTIVENESS

The Use of Control Groups. Before specific techniques are discussed, it should be pointed out that control groups are indispensible in determining the amount of improvement in job performance caused by a training program. Besco, Tiffen and King state the case for control groups by saying that, "To attribute changed attitudes . . . and improved job performance solely to a training program is impossible" unless the trained group is compared with "a control group that did not receive training."¹

The basic reason for this being true is that if a control group is not used, it is impossible to know whether the changes observed in the participants were produced by the program or merely brought about by the lapse of time.² Take, for example, a program designed to reduce the number of grievances. The grievance rate was .9 per year per hundred employees prior to the time the program commenced. After the participants completed the program and returned to their jobs, the grievance rate remained at .9 per hundred employees. Does this mean that the program was unsuccessful? No one can say so with a high degree of assurance. It might well be that physical or working conditions at the


plant changed to the extent that there were more genuine grounds for grievance and that, in the absence of training, the grievance rate might have risen to 1.5 per year per hundred employees. On the other hand, the program actually may have caused harm. Objective conditions may have changed for the better and had the supervisors not attended the program, the grievance rate may have fallen to .6 per year per hundred employees. In any event, no one really knows for certain whether the program attained its objectives, accomplished nothing, or did harm. This would not have occurred had a control group been used. By serving as a basis for comparison with the trained group, the control group would have made it possible to isolate changes induced by the lapse of time from those caused by the program.

While the use of a control group contributes to the elegance of any experiment used to measure the amount of information transmitted to the participants of a training program, its exclusion is of no great loss. The training program exists solely because it can transmit information in concentrated form and, for all practical purposes, constitutes a monopoly of information. Even though it is possible (though not likely) that factors outside the program may operate to furnish participants with information similar to the subject matter of the program, the possibility of their having a significant influence upon the total amount of information received by the participants is usually too remote to warrant going to the trouble of using a control group.

There is nothing to be gained by using a control group in connection with determining the acceptability of a program. A control group, as that term is used here, is not involved in the program; hence, it has no opinion about the acceptability of a program and cannot be used as a basis of comparison for a group which does have an opinion. It is well, however, to compare the
Techniques for Evaluating the Attainment of Program Objectives. Most prominent among the various techniques used to determine the extent to which programs have attained their objectives are the following:

1. Questioning the participants to find out their opinion about the program.
2. Giving attitude or achievement tests to the participants before beginning and after completing the program, with or without giving the same tests to control groups.
3. Comparing the participants' actual work performances before beginning the program with their performances after completing the program, with or without making the same comparisons with control groups.
4. Comparing ratings made by the participants' superiors or peers of the participants' performances before beginning and after completing the program, with or without making comparisons between similar ratings of control groups.

Examples taken from the literature of evaluation illustrating the use of each method will be described and the limitations of each method will be discussed in turn.

The most common method of evaluating programs is to question participants about their opinions of the program and to use the answers as an indication of the program's worth. The use of this method was reported by C. L. Van Sickle in connection with an evaluation of a University of Pittsburgh development program.

It was used along with other methods by Blansfield to evaluate a similar university program. Indeed it is so common, that in the literature of evaluation it is criticized more than it is reported.

Although evaluations by participants have at least one vocal defender—Kenneth Andrews says that "despite all the difficulties of involvement and subjectivity, the man undergoing an educational experience is the nearest to being an authority about its nature..." they have many more critics than supporters. Melvin Anshen neatly summarized the critics' point of view when he wrote that, "Popularity is no guarantee of utility. It may even have a contrary effect." Joseph Bailey confirms Anshen's comments by relating from his own experience incidents where students criticized those aspects of learning which contribute most to their learning. According to Bailey, the process of learning is a painful one which involves upsetting old and sometimes cherished patterns of thought in order to acquire new ones. The process is confusing and, in some instances, downright agonizing. Students who are undergoing, or who have undergone, this experience are quite likely to judge the value of a program by its difficulty and criticize that which helped them most.

These criticisms are well-taken. Learning is not easy and, John Dewey to the contrary, it is sometimes unpleasant, especially when assumptions and patterns of thought underlying much of our conduct are upset. Program

---


participants, because they are too close to the program and because they lack
the necessary training, cannot be relied upon to judge programs objectively.
Thus a valuable learning experience may be criticized by the participant, not
because it didn't teach him anything, but because it was uncomfortable. More­
over, a learning experience praised highly by the participant (because it is
pleasant and well presented) may leave the participant but little better off
than before because it has little or no actual bearing on the objectives of the
program.

Evaluations by participants simply cannot supply satisfactory answers to
the questions, "what did the program contribute to the participants' work per­
formance," or "how much did the program teach the participants." The observer
knows no more about how closely the program has approached these two criteria
of success after the participants have evaluated it, than he did before. How­
ever, he does learn something about the acceptability of the program. Accord­
ingly, although evaluations by participants are without value in determining
the degree to which the program has achieved the first two program criteria,
they are quite adequate for gauging attainment of acceptability, the third
criterion.

One rather common method used to evaluate the results of management devel­
opment programs is to test the participants before and after the program and
attribute the improvement in test scores to the program. Both attitude and
achievement tests have been used for this purpose either with or without con­
trol groups. When attitude tests are employed, it usually is assumed that
changes in attitudes can be extrapolated to changes in work performance. The
results of attitude tests, therefore, are not used to measure the amount of
information transmitted to the participants or the acceptability of the program;
rather changes in scores on attitude tests are interpreted as representing changes produced by the program in actual work performance.

Achievement tests based on the subject matter of the program usually have more modest aims. Most often they are used to measure the amount of learning done by the participants without attempting to touch upon how well or how often the learning is applied in work situations. Occasionally, however, achievement tests scores, like attitude test scores, are used as indicators of improvement produced by the program in actual work performance.

Evaluations made using the test method probably are better reported in the literature than any other types. The following are representative of the rather large number of test method evaluations written up in management development and training publications.

Raymond Katzell reported that a training program in human relations was evaluated by administering to trainees (supervisors) the "How Supervise" and "Intellectual Alertness" tests before beginning the program and an alternate form of the "How Supervise" test after completing the program. Katzell says that all participating groups scored significantly higher on the "How Supervise" test after completion of the program. Supervisors with low initial scores improved the most. New supervisors showed greater improvement than supervisors with more experience. A high degree of correlation was found to exist between "How Supervise" scores and "Intellectual Alertness" scores.8

Robert Wyland reported a similar evaluation. Fifteen foremen were given training in supervising. Form A of the "How Supervise" test was administered

---

to them before training began. After training was completed, Form B, an equivalent form of the "How Supervise" test, was administered. The participants raised their scores on the second test an average of 18% above their scores on the first test.9

A number of evaluations involving the administration of "before and after" tests to both participating and control groups have been reported. Mosel and Tsacaris describe one which was made of a training course attended by eighty-three commissioned and non-commissioned officers of the U.S. Air Force. The program consisted of forty hours of training in management techniques and human relations given over a six-week period. A control group consisting of forty-four men who were matched with the participants in terms of experience, age, rank, intelligence, and so forth, but who did not participate in the program, was established. Before the beginning of the program, the participating and control groups were given Form A of the "How Supervise" test. After completion of the program both groups were given Form B of the same test. Comparisons of differences in the improvement of scores over the period by the two groups were used as the basis for calculating the effects of training upon the participating group. The outcome of the experiment, according to the authors, was that the trained group made a small but significant gain in scores over the untrained control group.10

Ralph Canter describes an evaluation made of a human relations training program using somewhat similar techniques. Two groups, each consisting of

---


eighteen men, were established. Each group was matched in terms of age, mental ability, position, years of experience, and so forth. One group was used as a control group. The other participated in the training program. A battery of six attitude and achievement tests was given to both groups before the program began and after it was completed. Test results of the two groups were compared and the differences between them were attributed to the training received by the participating group.\textsuperscript{11}

One of the most elaborate evaluations involving the use of tests is one reported by Morris Viteles in \textit{Personnel Psychology}. The evaluation was made of the Bell Telephone humanities program conducted at the University of Pennsylvania. Three separate groups of executives attended the program. An additional two groups were used as control. The first control group was used as a basis of comparison for the first participating group. The second control group was used as a basis of comparison for the second and third participating groups. During a three-day period immediately preceding and following the program, each participating and control group was administered a wide variety of achievement and attitude tests. The results of the tests indicated that the participants had been inculcated with the attitudes and types of information originally established as objectives for the program.\textsuperscript{12}

One might ask what types of criteria of program success can the test method of evaluation be used to measure. Achievement tests obviously can measure what the participants have learned. If the immediate objective of the


program is to produce certain changes of attitudes in the participants, attitude tests can be relied upon in most instances to supply fairly reliable information about progress toward this objective.

However, some serious doubts exist about the adequacy of using achievement or attitude test results as indicators of the effect management development programs have had upon job performance. Paul Buchanan points out that the use of tests, "involves the assumption that changes in standing on the test ... are correlated with improvements on the job." Is this assumption correct? Not always. Buchanan says that, "changes on test scores occur without changes in job behavior." To the logical objections of Buchanan, Walter Mahler adds an empirical one. He writes that the results of two major studies indicate that changes in test scores did not reflect changes in job performance. Edwin Fleishman, who conducted one of the major studies referred to by Mahler, found that while test scores immediately following training went up, what management considered to be adequate job performance actually went down.

Evidence presented by Levine and Butler adds strength to this criticism. They found that two groups of supervisors taught the same subject matter by different methods scored approximately the same on tests. However, one group showed hardly any improvement in actual work performance. The conclusions of the authors are worth quoting. They read:

It is clear that group decision was more effective in reducing prejudicial ratings of these factory supervisors than was the formal lecture. This in itself is a significant finding. But what seems to be even more striking is the fact that the lecture method

13 Buchanan, p. 336.
15 Fleishman, pp. 205-222.
had practically no influence upon the discrepancies in ratings. It is generally assumed that once an individual or a group of individuals learn that they have been behaving in a socially undesirable way, they will immediately take steps to change, particularly if it is clear to these individuals that it is their responsibility to eliminate such errors. Our findings do not support such a notion. The acquisition of knowledge does not automatically lead to action. [emphasis added]

When Fleishman, Levine and Butler's findings are extended to the test method of evaluating management development programs (logic forces us to do so), one must conclude it is not unlikely that a program might do no more than teach participants how to supply the "right" answers to achievement or attitude tests. In such cases, test results would be completely misleading. They would indicate substantial improvements even though actual job performance remained the same or became worse.

The ease with which scores on attitude tests may be changed without producing comparable changes in behavior is demonstrated by an experiment conducted by Alexander Wesman in which it was found that when persons taking attitude tests were told to assume different roles, the test results were completely different. Speaking about the test results, he said, "If one saw these distributions without foreknowledge of how they were obtained, he could only conclude that they represented two quite different groups of people." Wesman's findings clearly show that there is a very real possibility that participants of a given program might well learn what attitude those conducting the program expect them to display; give the "right" answers to attitude tests; but still show absolutely no improvement on the job.


Other authors make objections based on the nature of objective-type attitude and achievement tests. The results of so-called objective tests, concludes Paul Diederich, may be quite misleading. In a study conducted by him, the lowest one-fifth of a class made spectacular gains over the course of a year while the upper one-fifth either made little gain or actually appeared to have lost ground. Similar results have been found in management development programs.

Diederich, however, demonstrated that test results do not furnish a true picture of the amount of learning done by the upper and lower one-fifth of the class. The first reason for this is that scores on objective tests depend upon both ability and chance. In tests administered before the class began, a higher-than-average proportion of persons who guessed wrong on questions will be found in the lowest fifth and a higher-than-average proportion of persons who guessed right will be found in the highest fifth. When the class is re-tested, the laws of probability produce a natural shift to the arithmetic mean—that is, of those persons in the lower fifth who guessed wrong on the first test, more will guess right and of those persons in the highest fifth who guessed right on the first test, more will guess wrong.

A second factor which distorts the meaning of objective test scores is that the choices in multiple-choice questions of the type found in typical objective tests are not equally obvious. A poor student who learns in the course

18Katzell, pp. 319-329, found that supervisors with the lowest initial scores on objective tests showed the greatest improvement. He concluded that such programs produce the greatest results in those who need the most help. Diederich's experiment yielded the same type of raw data. However, after he made statistical corrections for the so-called regression effect, (movement of scores toward the arithmetic mean in accordance with the laws of probability) he found that the highest fifth of the class improved more than the lowest fifth.
of a year to avoid manifest errors will show much more apparent improvement than a good student who can raise his grade only by learning to make very fine distinctions between alternatives which seem equally plausible.\footnote{Paul B. Diederich, "Pitfalls in the Measurement of Gains in Achievement," The School Review, LXIV (February 1956), 59-63. Diederich points out that it is much harder for a student to go from a score of 80 to one of 85 than it is for him to go from 30 to 60.}

The weight of evidence supports the use of achievement or attitude tests to measure changes produced by the program in the level of information or the attitudes of the participants.

Moreover, despite limitations and shortcomings, achievement or attitude tests may be used with some degree of success as indicators of improvement in job performance produced by development programs provided; (1) they are used in conjunction with control groups and (2) the tests have been empirically validated on criterion groups performing the same type of work as that which the program seeks to improve.\footnote{Although there is much in the literature about the validation of certain types of tests for the armed forces or for manual jobs of one sort or another, the author found only one report of the validation of a test designed to act as an indicator of performance for professional or managerial people. See Marvin L. Frederick, "Testing the Tests," The Journal of Accountancy, CIII (April 1957), 42-47.} However, in view of the fact that only a relatively small number of persons usually attends such programs, the work necessary to validate the tests properly is hardly worth the effort. Usually it is easier to use other methods better suited to evaluating the performance of small groups.

Actual Improvement in Job Performance. The ultimate test of the value of any management development program is the degree to which it improves actual job performance. As already mentioned, a program may be eminently acceptable;
it may even convey a good bit of information; but if it does not improve job performance, it is a resounding failure. We have already dealt with some of the techniques which indirectly measure more or less accurately improvements in job performance. Now we will consider those which dispense with indirect measurement by going directly to the job and determining to what extent formal training has improved work performance.

Typically the direct technique involves measurement of improvements in the performance of that type of work which the program was designed to improve. If the program is designed to improve sales technique, sales by the participants before and after attending the program are compared. If it is supposed to improve safety, accident rates are compared. If it is aimed at reducing turnover, then turnover rates are compared, and so forth.

An extensive evaluation of this type is reported by Wallace and Twichel. The program evaluated was a one-year training program conducted by Purdue University for insurance agents. One group of men attended the course. Another group which received no training was matched individually on the basis of sales age, marital status, work experience, and aptitude index rating with members of the experimental group, and used as a control. The performance of each group was judged according to so-called success criterion consisting of: (1) survival through the period, (2) average sales of $15,000 per month, or (3) promotion. Findings of the study indicate that the program materially assisted the participants in achieving the "success criterion." Of the Purdue group, 41% met the criterion but only 22% of the control group did so. The success of members of the control group depended strictly upon their production record prior to the time training began. The success of the experimental group, however, did not depend upon pre-training production records. This shows that the program made
improvements which, in its absence would not have been made. The authors concluded that the major effect of the training program was to raise production levels of men who otherwise would have failed.21

Peter Castle reports a very complex method of evaluation used to determine the effectiveness of two programs and two types of classroom presentation (lecture and discussion). The programs consisted of classes in human relations. That part of the evaluation used to determine the effectiveness of the programs involved using trained observers to grade and compare the pre-program and post-program performances of participants in a human relations role-playing situation with those of a non-participating control group. The role playing situation, according to Castle, possessed "natural validity" inasmuch as it presented for solution the same type of problem that would be found on the job. In effect, it constituted a sample taken out of the actual work situation. Comparison of performances by the two groups revealed that the trained group, in contrast to the untrained group, improved markedly over its pre-training performance.22

There is not much question that comparing changes in the work performance of participants with those of a non-participating control group is a method of evaluation which will give an adequate answer to the ultimate question—did the program produce results. On this point, it is without peer. However, it lacks the ability to discriminate between the criteria of acceptability and communication of information. While it may show whether or not the program was a


success, it cannot show the relative importance of the roles played by acceptability and transmission of information in making it a success. Likewise, if the program is a failure, it is impossible to tell how much poor acceptability or failure to communicate relevant information contributed to the failure. Accordingly, while this method has been found to be an excellent one for determining the over-all success of a program, it is not up to the task of determining the relative weight of the contributions made by two important factors, acceptability and transmission of information, to the success or failure of the program.

**Ratings by Superiors, Peers or Subordinates.** Frequently, ratings made by superiors, peers or subordinates of the work performance of participants in a program are used as the standard by which a program is deemed a success or failure. Usually the ratings consist of replies by supervisors, peers or subordinates to questions asked about what changes have been observed in the participants' work behavior since completion of the program. Reports of favorable changes are interpreted to mean that the program has achieved its objective.

A typical evaluation of this type is reported by Paul Buchanan. The superiors and subordinates of the participants were given a questionnaire and asked to rate the participants on specific work habits enumerated in the questionnaire. No formal rating was made prior to the program although it was assumed that those making the ratings would be able to recall the level of the participants' work performance prior to attending the program, and no control group was used. On the basis of returns of the questionnaires, the author concluded that two-thirds of the participants showed desirable modifications in
their work habits following the program. 23

Daniel Goodacre reports that a modified version of this technique was used to evaluate a leadership training program attended by four hundred persons. Members of the participant group were given achievement tests and rated by their superiors before and after attending the program. A control group of four hundred persons was tested and rated at the same times. The evaluation indicated that as a result of the program, the experimental group improved in several respects over the control group. 24

The rating technique is a method for determining the effectiveness of work performance by employees. When it is used to evaluate a training program, it serves as an indicator of the degree of improvement in actual work performance caused by the program. As pointed out earlier in the discussion of the use of the technique of directly measuring actual work performance, accurate measures or indicators of work performance will tell the observer whether the program generally was a success. However, such techniques are unable to discriminate between the acceptability criterion and the transmission of information criterion. The rating technique suffers from the same defect. It cannot be used to measure the acceptability of a program or the amount of information transmitted by the program to the participants. If it is to be used at all, it must be used as a general indicator of the improvement in work performance wrought by the training program and, as is true of other techniques used for this purpose, it must be used in conjunction with control groups.


24Goodacre, pp. 534-538.
The question may be asked whether the rating technique, even when used with control groups, will produce reliable information about the effectiveness of a training program. Basically, the rating technique consists of a judgment passed, without the benefit of quantitative measurements, by someone upon the adequacy of the work of a given employee. It does not involve a direct measurement of work performance; rather it is expressed opinion or educated guess, depending upon the point of view, about the effectiveness of an employee. By its very nature, it involves the use of qualitative standards and like all methods involving qualitative standards, it is filled with difficulties. Thorndike summarizes the major difficulties found in ratings by saying, "The standard varies from rater to rater from time to time and from place to place," and that the results may be further biased should the rater be "prejudiced in favor of some one of the particular training programs that are being compared." He concludes on this note of warning; "in investigations of particular experimental or training procedures the use of ratings as a criterion must be viewed with critical suspicion."25

These are serious objections which, moreover, are generally admitted to be correct. However, this is not to say that the technique is worthless. While the data it yields may not be sufficiently exact to make fine distinctions between degrees of program success, they are adequate to show, in a general way, whether the program has succeeded or failed.

Concluding Statement. Four techniques used to evaluate the effectiveness of training programs have been described. The adequacy and limitations of each technique relative to measuring attainment of the three basic program criteria

(acceptability, communication of information and improvement in work performance) have been discussed in considerable detail. Some of the techniques serve rather adequately in measuring attainment of one or two of the basic criteria. None, however, is suitable for measuring attainment of all three basic criteria. Some, even where applicable, are not as effective as others.

The purpose of this chapter was to establish the factual and argumentative bases for ranking the effectiveness of each technique in measuring the attainment of the three basic program criteria. It is earnestly hoped that that objective has been achieved. However, since other things must be accomplished before an actual ranking of evaluational techniques in terms of their relative effectiveness is made, that task will be reserved to Chapter VI.
CHAPTER V

TECHNIQUES OF EVALUATING TEACHING METHODS

Purposes and Nature. In the search to find new ways to improve the quality of formal programs and secure the greatest benefits from the training budget, teaching methods in use should, as a matter of course, be subjected to careful analysis. Those which prove to be more effective or less costly should be adopted in favor of those shown to be less effective or more costly.

It should be remembered, however, that there is no such thing as a teaching method which is superior to all others in all learning situations. Therefore, it is not enough that a given method has been evaluated once. Whenever a new program is started, the search for the best teaching method for that particular program must begin anew. The purpose of the teaching method evaluation, then, is not to find the best teaching method per se. It is to find the best method for the particular type of program under consideration. The evaluation of teaching methods, thus considered, is not a sporadic thing; rather it is, or should be, part of a continuous process.

Criterion. What standard determines the best teaching method in a given context? There is not too much disagreement on the standard. Most authorities in the field of education agree pretty much that the effectiveness of teaching methods can be best judged by the amount of information conveyed to and retained by the participants.¹

Evaluational Methods, General. Teaching methods may be evaluated in one of two ways. First, a given method may, without reference to other methods, be compared with an established standard. If it meets the standard it is adequate; otherwise not. Secondly, a given method used to teach certain subject matter may be compared with another method used to teach the same subject matter. This type of evaluation, because it goes beyond the first by showing not only which methods are adequate, but also which one of a number of methods is best, is considered superior to the first.

There are four evaluational techniques in general use. These are; (1) achievement and other test scores, (2) ratings by instructors, (3) ratings by participants, and (4) ratings by outside observers. In practice, usually only the first technique and, sometimes the fourth technique are used as a basis for comparing the effectiveness of one teaching method with another. All of the techniques, however, may be used with varying degrees of success to determine whether a given method has reached a pre-determined standard of adequacy.

Achievement and Other Objective Test Scores. Using achievement or other test scores to evaluate the effectiveness of teaching methods has long been recognized as one of the more adequate evaluational techniques. They may be used with equal facility to determine whether a given method has attained a pre-established standard of adequacy or which of two or more methods is best suited for teaching a given subject matter. Of all the evaluational techniques, it is the one used most frequently for comparisons between teaching methods.

Rating By Instructors. Using ratings made by instructors as the basis for evaluating the effectiveness of teaching methods is a somewhat less desirable technique that the one just discussed. Although it will disclose instances
where teaching methods are very inadequate, it still has the same difficulties usually associated with rating techniques. Generally speaking, no serious effort is made to see that all instructors understand the rating procedure alike. As a result, each is apt to have his own opinion about what is needed for a given method to reach a pre-established standard of adequacy. If the relative effectiveness of two or more methods is to be determined by this technique, the difficulties are multiplied. The comparative ratings given by a number of instructors for the methods may amount to little more than an aggregation of personal opinions based on private understandings of standards of adequacy.

Ratings By Trained Observers. Ratings of teaching methods by observers trained in the use of standardized rating guides is a somewhat better technique. Usually there is adequate consistency (rater reliability) between ratings made of the same technique by different observers. This makes it possible to compare two or more teaching methods without observer bias gravely influencing the results. Nevertheless, the technique has its limitations. It does not purport to measure directly what participants subjected to a given teaching method have learned. It merely passes judgment (admittedly an "expert" one) upon what appears to have been learned. It is, because of its consistency, superior to ratings made by instructors. However, it is not as effective as methods which directly measure participant learning.

Rating by Participant. Having program participants rate the value of different teaching methods is a technique of evaluation which has proven popular in some circles. The way in which it is typically used may be illustrated by discussing an evaluation reported by Francis J. DiVesta. The evaluation was made of teaching methods used in a lengthy development program for army officers. The methods used were lectures, demonstrations, discussions, seminars,
individual training and staff exercises. The participants were asked to (1) list the most valuable and least valuable technique (forced choice method) and give reasons for their choice and (2) describe the amount of productive time experienced with each method. The participants rated lectures, staff exercises, and seminars most highly. The author stated that the ratings were not conclusive as to the value of the methods since the actual learning produced by each had not been determined. He felt, however, that there was some indeterminate relationship between ratings of the methods by the participants and the actual effectiveness of the methods.²

There are many objections to using ratings by participants for evaluational purposes. The participant is not a trained observer. Consequently, it is more likely than not that his judgment will be based on private standards. He may, for example, find that some methods are more entertaining than others and although he learns no more, (and sometimes less) under one method than under another, he may rate the former higher because he enjoyed it more.

Evidence about bias in ratings of teaching methods by students is mixed. Marsh, Burgess and Smith found that students tend to rate methods in terms of their classroom performance. The better a student performs, the higher he will rate the method used.³ On the other hand, according to an article appearing in Higher Education, the ratings of instructors by students show no systematic bias—they were consistent regardless of the status of the student (graduate or


³Joseph E. Marsh, George G. Burgess and Paul N. Smith, "Student Achievement as a Measure of Instructor Effectiveness," The Journal of Educational Psychology, XLVII (February 1956), 79-88.
undergraduate), whether the classes were large or small, and what grades the students received. It is noteworthy, however, that there was little agreement between ratings made by students and those made by the instructors' peers and superiors.

All things considered, ratings by participants probably is a more accurate gauge of how well the participants liked the technique rather than how effective it was. Of course, if a given teaching method proves to be universally unpopular, it probably should be abandoned. To this extent ratings by participants are useful. However, because of the widely varying standards of the participants, it simply is impossible to use the method to determine the relative effectiveness of various teaching methods. For this reason and because of the possibility of participants mistaking that which is entertaining for that which is effective, ratings by participants must be deemed the least satisfactory of the devices used generally for evaluating the effectiveness of teaching methods.

Evaluations, Comparisons of Methods. Most of the reports on evaluations of teaching found in educational journals (relatively few are found in business of personnel publications) are of the "comparison of methods" type. Typically, they involve selecting two matched groups, testing them before the beginning of the training program to determine their existing knowledge of the subject to be studied; using different teaching methods to instruct each group in the same subject matter; and testing each group again after completion of the training program. The method used with the group showing the greater improvement

in test scores is adjudged to be the better of the two. In some instances, this technique is refined further by retesting both groups following a lapse of time after completion of the training program. The difference between those test scores and scores made on tests given immediately after completion of the program is interpreted as representing the effect each teaching method had upon the participants' retention of information.

**Examples.** It would be worthwhile to digress for a moment to review some of the studies which evaluate by comparing the effectiveness of two or more teaching methods. Not only are they intrinsically interesting, but they illustrate the steps which should be followed in order to insure precision in the evaluation.

The literature of education contains reports of comparative evaluations made of almost every type of major teaching method from audio-visual aides to the case study technique. Ralph Bentley, for example, relates in detail two experiments made by him to determine the contributions made to learning by audio-visual aides. Classes in soil conservation were subjects of the first experiment. Eleven experimental groups and eleven control groups of equal size were used in the study. Each experimental group was paired with a control group and the two were placed under a single teacher. Each group was given a mental ability test and soil conservation achievement test before the class began. Texts containing many photographs and illustrations were assigned to both groups. Both were taught by the usual lecture method. In addition, the experimental groups were shown a series of motion pictures dealing with soil conservation. After the class was completed, a soil conservation test, equivalent to the first test, was administered to each group.
Improvement in post-class test results indicated that the experimental groups learned no more than the control groups.

The second experiment involved classes in "permanent pasture production." The evaluational technique used was the same as that used in the first experiment except that a larger number of experimental and control groups were used. Both groups were taught by the lecture method. However, in contra-distinction to the first experiment, the text used by the classes contained no photographs or illustrations. A series of films and slides on pasture production were shown to the experimental groups. The post-class achievement test scores indicated that the experimental groups learned significantly more than the control groups.

The results of the two experiments led Bentley to conclude that audio-visual aids are expensive luxuries when the text used contains an adequate number of photographs and illustrations. In fact, he says that in such circumstances, audio-visual aids are repetitious and may, by taking up the students' time, prevent them from learning as much as they might have had the time been spent studying the text. However, he goes on to say that when available texts do not contain photographs and illustrations, audio-visual aids are well worth using.5

In an evaluation made by J. Darrell Bernard, the lecture-demonstration and the problem-solving methods of teaching were compared. The subject of the experiment were college-level general biology classes. Each teaching method was evaluated in terms of its effect upon the students' (1) recall of

information, (2) understandings of generalizations, (3) problem solving ability, and (4) "scientific attitudes." Three classes totaling 145 students were taught by the problem-solving method. Three classes totaling 135 students were taught by the lecture-demonstration method. Prior to the time classes began, all groups of students were equated on the basis of psychological tests and biology achievement tests.

Barnard designed four tests to measure attainment of the four objectives cited above. One experimental group and one control group were given the battery before beginning and after completing the course. A second experimental group and control group were given the battery before beginning and after completing the first half of the course. The remaining experimental and control groups were given the battery before beginning and after completing the second half of the course. In all three instances, test results indicated that neither of the two methods produced a significant difference in the recall of information or the understanding of generalizations. However, students taught by the problem-solving method improved significantly in their ability to solve problems. Their general "scientific attitudes" also showed improvement.  

It is not safe, however, to conclude that because using the problem-solving method to teach biology classes improves the ability to solve biological problems, it will do the same for all types of subjects. Different


results from those of Barnard's were achieved in an experiment reported by Leonard A. Ostlund. The experiment involved teaching economics to an experimental group by the case-discussion method—one similar to the problem-solving method. A control group, matched individually with members of the experimental group, was taught economics by the lecture method. Before the class began, the members of both groups were each given problems to solve. In addition, each group was assigned a problem to solve collectively. The attempt at collective solution was unsuccessful for both. However, individual solutions were made by members of both groups and judged by a panel of experts. After the course was completed, the process was repeated. Again neither group was able to solve the problem assigned to it for collective solution. The panel of judges found solutions proposed by individuals of each group to be comparable value. Ostland said that the only noticeable difference between the two groups was that the experimental group showed improved ability to work as a group. However, the improvement was not sufficient to produce a collective decision on the group problem.  

Harry Ruja reports an evaluation made of the effectiveness of the lecture method and the discussion method in teaching psychology and philosophy classes. Students in the philosophy and psychology classes were divided into two groups. One group of philosophy and psychology students were taught by the lecture method. The remaining group of philosophy and psychology students were taught by the discussion method. All groups, among other things, were given achievement tests before beginning and after completing the classes. Ruja found that while philosophy students did equally well under either method,
students in psychology did significantly better when taught by the lecture method.

Luther Colyer describes an evaluation in which he compared the lecture method with the "principle-unit" method of teaching biology classes. The "principle-unit" method consists of organizing class work around certain important generalizations of biological science, and having students select the materials needed to study and illustrate the principles. Students from three colleges participated in the study. The participants were separated into twelve groups. Six groups were taught by the lecture method and the remaining six by the "principle-unit" method. Prior to the time classes began, all groups were tested for intelligence, knowledge of biology, biological aptitude, and general scholastic achievement. Following completion of the class each group was retested.

The results of the experiment did not indicate marked superiority on the


Also see George Hunsberges, "An Experiment in Educational Methods," Journal of Business Education, XXXIII (April 1958), 283-284. Hunsberges made an experiment similar to that of Ruja's and found that so far as scores on achievement tests are concerned, there was no noticeable difference in the effectiveness of the two methods.

It is noteworthy that Ruja asked each student to evaluate the method by which he was taught. The discussion method had led to greater acquaintance among the students (this was measured by Ruja) and to considerable socialization outside of class. The philosophy students rated the discussion method considerably higher than the lecture method, despite the fact that, objectively speaking, they did as well under one method as the other. Psychology students rated both methods the same despite the fact that they learned significantly more under the lecture method. This points up one of the dangers of using student evaluations as a basis for determining the effectiveness of a technique; namely, that the student frequently mistakes that which is pleasant for that which is effective.
part of either method. The "principle-unit" groups raised their scores on the biological aptitude tests somewhat above (but not significantly above) those of the lecture groups. However, there was little, if any, difference between the two groups in knowledge of biology and comprehension and application of biological principles. ⑨

One of the more interesting evaluations found in educational literature was devised and carried out by John Ward. Ward compared learning between a class taught by the "group-study" method and another taught by the lecture-demonstration method. In the group-study method the students assume many of the functions of the instructor in the lecture-demonstrative method. They are expected to define class objectives, decide upon classroom activities and evaluate their own progress. The instructor has a passive role. For the most part he merely serves as a source of information for the class.

The students of two general science courses participated in the experiment. Both classes were taught by Ward. Before classes began the students of each were equalized on the basis of American College Board test scores and general science achievement tests. Immediately after the class was finished, each class was tested to determine its knowledge of the subject matter and understanding of the application of general scientific principles. Later each group was retested to determine its retention of knowledge of subject matter and principles.

The results were quite interesting. That part of the group-study class whose scores on the first tests ranked in the upper on-fourth of the class

⑨Luther Colyer, "Comparison of Two Methods of Teaching Biology at the College Level," Science Education, XLIV (February 1960), 52-58.
showed much more improvement than the comparable one-fourth of the lecture-demonstration class. The results for the remainder of the class were reversed. The lower three-fourths of the group-study class did not progress as well as the lower three-fourths of the lecture-demonstration class. 10

The foregoing reports demonstrate what may be accomplished when evaluations are carefully thought out and executed. If one method of teaching is superior to another the evaluation points up that fact. If they are of comparable value, that, too, is shown. Equally important is the fact that the care with which the evaluations were made insures that their findings can be relied upon as the basis for practical decisions.

Incidentally, the evaluations just discussed make one other important point. By showing that the effectiveness of any given teaching method may depend upon the subject matter being taught or the level of achievement of the participating group, they demonstrate that no single method should be considered, a priori, the best. Rather, in the absence of persuasive evidence to the contrary, the effectiveness of any given method should be considered limited to the subject matter and level of achievement present in the original evaluational situations. This re-emphasizes the importance of re-evaluating the same teaching method when it is used to teach different subject matter or students at different levels of achievement.

Concluding Statement. It has been shown that inasmuch as the objective of a teaching method is to produce learning in the student, the proper standard

for measuring its effectiveness is the amount of information learned by the student. It has been further shown that evaluations, in general, are most effective when used to determine the relative merits of two or more types of teaching methods. Examples have been given of the techniques used generally to evaluate the effectiveness of teaching methods and their usefulness, limitations and relative efficiency have been discussed. The actual ranking of the relative effectiveness of evaluational techniques, however, is reserved for the next chapter.
CHAPTER VI

RATING PROGRAM EVALUATION PLANS

Most of the preceding portions of this paper have been devoted to identifying which features are essential to good program evaluation plans. At this point we will summarize briefly the points which have been made thus far.

Each evaluational plan has a procedural aspect and a technical aspect. The procedural aspect refers to the scope of the plan and the uses made of evaluational data. The technical aspect relates to the adequacy of evaluational methodology. From the procedural standpoint an evaluational plan should (1) evaluate all types of programs conducted or participated in by the company; (2) evaluate teaching methods used in programs conducted by the company; (3) use evaluation results as the primary basis for deciding whether a given program or teaching method will be discontinued or modified. From the technical standpoint, methods for evaluating programs should be capable of measuring effectively the three criteria--improvement in work performance, acceptability and transmission of information--discussed in Chapter III. In the evaluation of teaching methods, the techniques used must, as a minimum, be capable of determining the amount of learning done by the participant group. Preferably, because of the relative nature of teaching method effectiveness, they should be used in such a way as to permit comparison between two or more teaching methods. When they are so used, the number of variables which might have a bearing upon the outcome of the evaluation should be kept as low as possible by using matched groups and other techniques to equalize experimental conditions.

One of the stated purposes of this paper is to determine the effectiveness
of plans devised to evaluate management development programs. The remainder of this chapter will be spent in developing a system by which the procedural and technical aspects of such plans can be rated.

**Rating Procedural Effectiveness.** It will be recalled that an evaluational plan which is completely effective from the standpoint of procedure, (1) evaluates all programs in which the company is involved, (2) evaluates teaching methods that are within the control of the company and (3) serves as the basis for decisions to modify or abandon programs or teaching methods. The procedural effectiveness of evaluational plans will be rated on a point system built around these three procedural steps.

Devising a rating system presents difficulties. It is possible for one plan, to cover completely all phases of the company's training programs even though it is not nearly as extensive as that of another company. Take, for example, a company which participates only in short *ad hoc* "outside" management development programs. If it evaluates the programs and uses the evaluational data as the basis for decisions about the future of the programs, its evaluational plan is procedurally complete viz., it covers all phases of the company's rather limited training endeavors. On the other hand, a company which, in addition to sending men to outside programs, conducts "inside" management development programs would have to evaluate both types of programs as well as teaching methods used in connection with "inside" programs and use evaluational data as the basis for decisions about the future of programs and teaching methods in order for its evaluational plan to be procedurally complete.

In order to take this factor into account, a negative system of scoring must be devised. Each evaluational plan will be assigned an arbitrary score from which deductions are made when the company under consideration operates in
an area of training not covered by the evaluational plan. The score remaining after all deductions have been made, will represent the procedural effectiveness of the evaluational plan.

The score assigned to each evaluational plan before deductions is sixty. Failure to evaluate programs and teaching methods may lead to a maximum deduction of thirty points. Failure to evaluate repeated programs and teaching methods or after having evaluated them, to use evaluational data as the basis for decisions about abandonment or modification of programs or methods may lead to the deduction of an additional thirty points.

The reader will note that a one-to-one ratio exists between failure to evaluate programs and teaching methods and failure to use evaluational data as the basis for decisions about the future of programs or teaching methods. This ratio was established because, in the author's opinion, an evaluation which is not put to use might as well not have been made.

Procedural Evaluation Program Coverage. For the purpose of this study, management development programs were divided into two general groups—inside and outside programs. An inside program is defined as being a "management development program conducted for managerial employees by respondent company." An outside program is defined as being, "management development program attended by managerial employees of respondent company, but conducted by organizations outside the company, i.e., AMA seminars, university programs, etc., other than those leading to a formal degree."

Inside and outside programs were broken down into four categories based on (1) the amount of time consumed by the program and (2) whether or not they were conducted more than once. Thus, the total maximum number of program
types is eight. These are; inside and outside ad hoc programs\(^1\) consuming less than twenty hours of the participants' time; inside and outside ad hoc programs consuming twenty hours or more of the participants' time; inside and outside repeated programs\(^2\) consuming less than twenty hours of the participants' time; and inside and outside repeated programs consuming twenty hours or more of the participants' time.

Failure to evaluate certain types of the programs described above is not as serious as failure to evaluate other types. Evaluating repeated programs is considered more important than evaluating ad hoc programs inasmuch as evaluation of repeated programs will not only tell whether the programs have been a success, but whether they should be abandoned or changed. It is more important to evaluate long programs than short ones inasmuch as the former involve greater expenditures of time and money.

After taking these factors into account, it was decided that deductions should be made as follows when a company conducts or participates in, but fails to evaluate, any one of the eight types of management development programs just enumerated.

---

\(^1\)The term "ad hoc program" is defined "special purpose program which is given only once."

\(^2\)The term "repeated program" is defined "program attended by successive groups of participants. Differs from ad hoc program in that it is conducted more than once."
TABLE I
RAW NEGATIVE SCORES: EVALUATION OF PROGRAMS

<table>
<thead>
<tr>
<th>Subject Present But Not Evaluated</th>
<th>Negative Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outside, Short Ad Hoc Program</td>
<td>1</td>
</tr>
<tr>
<td>Inside, Short Ad Hoc Program</td>
<td>1</td>
</tr>
<tr>
<td>Outside, Long Ad Hoc Program</td>
<td>2</td>
</tr>
<tr>
<td>Inside, Long Ad Hoc Program</td>
<td>2</td>
</tr>
<tr>
<td>Outside, Short Repeated Program</td>
<td>3</td>
</tr>
<tr>
<td>Inside, Short Repeated Program</td>
<td>3</td>
</tr>
<tr>
<td>Outside, Long Repeated Program</td>
<td>4</td>
</tr>
<tr>
<td>Inside, Long Repeated Program</td>
<td>4</td>
</tr>
<tr>
<td>Teaching Methods</td>
<td>10</td>
</tr>
</tbody>
</table>

In instances where the company under consideration has all eight types of programs, deductions for failure to evaluate programs or teaching methods will be made exactly as shown in Table I. However, when the company has less than eight types of programs, failure to evaluate a given program becomes proportionately more serious. If the raw negative scores were used directly, this fact would not be reflected. Therefore, formulas were devised to overcome this problem by making deductions proportionate to the total number of programs conducted or participated in by a given company.

Accordingly, deductions for companies with outside programs only will be computed as follows:

\[
\text{Summation Negative Scores, Programs Not Evaluated} \times 30 = \text{Deduction.}
\]

\[
\text{Summation Negative Scores, All Programs Present}
\]

Where less than eight types of programs are involved, but where an admixture of inside and outside programs are found, failure to evaluate teaching methods...
will result in a deduction of ten points and remaining deductions will be computed thus:

\[
\text{Summation Negative Scores, Programs Not Evaluated} \times 20 = \text{Deduction.}
\]
\[
\text{Summation Negative Scores, All Programs Present}
\]

The author is well aware that the foregoing scheme is far from perfect. The ratios established between various negative scores are based solely on the author's judgment. Important facts such as the relative amounts expended for each type of training program could not enter into the judgment because they were not available. However, it should be remembered that although the rating system is an approximate creature, the data it will be called upon to evaluate are also approximate. Accordingly, the rating system is probably adequate for the type of data to which it is to be applied.

**Evaluation Data Used As Basis For Decisions.** The final point by which the procedural completeness of an evaluational plan will be judged is whether evaluational data is used as the basis for decisions about changing or abandoning programs and teaching methods. Deductions in this area cannot be made for failure to make and use evaluations of ad hoc programs--by definition, they are given only once; hence it is logically impossible for them to be changed or abandoned on the basis of evaluational data. However, when repeated programs or teaching methods under the control of the company are not evaluated\(^3\) or when they are evaluated but the evaluational data is not used as the

---

\(^3\)Because of the system of negative scoring, it is necessary to make deductions in instances where repeated programs are not evaluated. Otherwise, a company which failed to evaluate a repeated program would have the same overall procedural score as one which evaluated the program but failed to use the results as the basis for administrative decisions.
basis for decisions about changing or abandoning programs or teaching methods, deductions will be made. A deduction system, adjusted to reflect the fewer number of possible elements, similar to the one discussed in the preceding system has been constructed for this purpose. The raw negative scores for the system are shown in Table II.

TABLE II

RAW NEGATIVE SCORES: USE OF EVALUATIONAL DATA

<table>
<thead>
<tr>
<th>Subject Present, But Not Evaluated, or Evaluational Data Not Used</th>
<th>Negative Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outside, Short, Repeated Program</td>
<td>4.3</td>
</tr>
<tr>
<td>Inside, Short, Repeated Program</td>
<td>4.3</td>
</tr>
<tr>
<td>Outside, Long, Repeated Program</td>
<td>5.7</td>
</tr>
<tr>
<td>Inside, Long, Repeated Program</td>
<td>5.7</td>
</tr>
<tr>
<td>Teaching Methods</td>
<td>10.0</td>
</tr>
</tbody>
</table>

Each deduction made for program evaluations enumerated in the table has been computed as a fraction of twenty (the maximum deduction made when none of the evaluations of programs are used) in the same ratio as the deduction made for a given type of program in Table I bore to total possible deductions for failure to evaluate repeated programs.

In cases where the company under consideration has all four types of repeated programs, deductions for failure to evaluate them or teaching methods or to use evaluational data as the basis for important administrative decisions will be made as shown by Table II. However, when the company has fewer than four types of repeated programs, deductions will be computed in accordance with formulas devised to keep deductions proportionate to the number of elements present in this area for any given company. When the company has outside repeated programs only, deductions will be computed as follows:
Summation Negative Scores, Failure to Make or Use Evaluations \( x 30 = \text{Deduction.} \)

Summation Negative Scores, All Repeated Programs Present

Where less than four types of repeated programs are involved, but where an admixture of inside and outside repeated programs are found, failure to evaluate teaching methods or failure to use the results of evaluations made of teaching methods will result in a deduction of ten points. Remaining deductions will be computed in accordance with the following:

\[
\text{Summation Negative Scores, Failure to Make or Use Evaluations} \times 20 = \text{Deduction.}
\]

**Technical Effectiveness.** Companies in the sample will be rated separately on the effectiveness of techniques used by them to evaluate; (1) programs and (2) teaching methods. Evaluational techniques used to determine the effectiveness of programs will be rated in terms of their capacity to measure the attainment of three criteria—(1) acceptability, (2) communication of information and (3) improvement in work performance. Techniques used to evaluate teaching methods will be rated according to their effectiveness in measuring gains in participant learning.

The rating system by which the technical effectiveness of evaluational methods will be judged differs from those used to rate procedural effectiveness in that positive scores rather than deductions from a fixed score are used. This is made possible because certain types of methods must be included in the system of evaluational methods in order for it to be technically effective. This was not true in the case of rating the procedural effectiveness of evaluational plans. Any type of training program could be present or absent from the procedural system without disturbing its effectiveness.

The following table shows the scores which are assigned to various types
of evaluational techniques used generally to measure achievement of each of the three major program criteria. Evaluational techniques used in connection with inside and outside programs will be rated separately. A maximum score of nine can be achieved for evaluational techniques used with either type of program.
### TABLE III

EVALUATIONAL TECHNIQUES: APPLICATION AND EFFECTIVENESS

<table>
<thead>
<tr>
<th>Technique</th>
<th>Criteria - Score&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Acceptability</th>
<th>Communication of Information</th>
<th>Improvement in Work Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questioning Participants</td>
<td></td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Testing Participants, no control group</td>
<td></td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Testing Participants, Control group</td>
<td></td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Measurement of Work Performance, No Control Group</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Measurement of Work Performance, Control Group</td>
<td></td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Rating of Work Performance, No Control Group</td>
<td></td>
<td>0</td>
<td>0</td>
<td>.5</td>
</tr>
<tr>
<td>Rating of Work Performance, Control Group</td>
<td></td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

<sup>1</sup> Where companies employ two techniques, either of which can be used to measure a given criterion, only the technique with the higher value for that criterion will be counted.
The following table shows the scores which are assigned to various types of evaluational techniques used to measure the effectiveness of teaching methods. A maximum score of nine may be achieved. Where more than one evaluational technique is used, only the one giving the highest score will be used in the rating.

TABLE IV

TEACHING METHODS EVALUATIONAL TECHNIQUES

<table>
<thead>
<tr>
<th>Technique</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation by Student, No Comparisons</td>
<td>1</td>
</tr>
<tr>
<td>Evaluation by Instructor, No Comparisons</td>
<td>2</td>
</tr>
<tr>
<td>Evaluation by Observer, No Comparisons</td>
<td>3</td>
</tr>
<tr>
<td>Achievement Tests, No Comparisons</td>
<td>4</td>
</tr>
<tr>
<td>Controlled Comparisons of all Types</td>
<td>9</td>
</tr>
</tbody>
</table>

\(^1\)Least effective to most effective represented by range 0-9.

Combined Scores. For purposes of comparison, it is desirable that the procedural and technical scores or different types of technical scores for a given company be combined and compared with those of other companies. This presents several problems. First, a company which has and evaluates inside programs, outside programs and teaching methods using the most effective techniques will have a total technical score of twenty-seven, thirty-three points less than the maximum possible procedural score. Since the procedural and technical aspects of evaluational systems are considered of equal importance, it is obvious that technical scores must be converted so that a maximum score of sixty
is possible. When this is done, both scores will have equal weight in the combined score.

Secondly, the use of a positive scoring system for technical evaluational methods creates an additional problem. One will recall that no matter how few types of programs a given company has, it may achieve the maximum procedural score if it meets certain standards. However, this is not the case with the technical scoring system. A company could use the most effective techniques to evaluate, say, outside programs (the only ones it has), and receive a lower over-all technical score than a company which uses indifferent techniques to evaluate a wide variety of inside and outside programs. This presents no problem so long as only technical scores for given categories of programs are being compared. However, if the scores are to be combined for any purpose, an adjustment must be made which eliminates this source of distortion.

After these factors were considered at some length, it was decided that the formula set forth below was the one best suited for converting raw technical scores, into scores which could be used either to arrive at combined technical scores or combined procedural-technical scores.

\[
\frac{\text{ATS}}{\text{AC}} \times 7 \times 2.2 = \text{CTS}
\]

"ATS" stands for the total of actual technical scores; "AC" stands for the total number of categories in which technical scores (including zero) were found; the number seven equals the total number of possible technical categories; and the number 2.2 is the conversion factor needed to change a maximum technical score to thirty. In order to arrive at a combined technical-procedural score for a given company, the converted technical score is merely added to the actual procedural score.
One major assumption is made in the equation; namely, that if the company under consideration were to extend its evaluation plan to areas not now covered, the level of effectiveness for methods used to evaluate the new areas would be no higher or lower than the level of those it presently uses.

The assumption is not too far wrong inasmuch as those companies which evaluate both inside and outside programs usually employ comparable methods of evaluation for both and, consequently, receive approximately comparable technical scores for both.

Method of Using Rating Systems. One of the major problems involved in the use of questionnaires to gather information is the fact that many which contain valuable information were returned without having been entirely completed. To be absolutely correct from the methodological standpoint, those which were not fully completed should be discarded. However, if this is done much information would be lost. In order to save as much as possible, it has been decided to use information contained in incomplete questionnaires wherever possible. Accordingly, ratings will be made of individual companies wherever there is complete information in any given area even though questionnaires from the same companies might be incomplete in other respects. This means that varying numbers of companies will be included in specific categories of the rating systems just described. However, where scores are combined or correlations are made which require companies to supply answers in two or more areas, only questionnaires containing those answers will be used.

The preliminary work of constructing evaluational tools has now been completed and we shall turn to the study proper.
Returns. Questionnaires were sent to one hundred corporations. Sixty-nine responded by returning questionnaire. Of these, fifty-seven (82.6%) have formal management development programs of one type or another. The remaining twelve companies (17.4%) are without formal programs.

Outside Management Development Programs. Fifty-two companies representing 91.2% of all companies with formal management development programs, reported that they participate in outside programs. Of this number, fifty (96.2%) specified the types of outside programs in which they participate. Table V below summarizes this information.

**TABLE V**

**TYPES OF OUTSIDE PROGRAMS REPORTED BY FIFTY COMPANIES**

<table>
<thead>
<tr>
<th>Type of Programs</th>
<th>Companies Reporting Outside Programs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
</tr>
<tr>
<td>Short Ad Hoc Program</td>
<td>38</td>
</tr>
<tr>
<td>Short Repeated Program</td>
<td>31</td>
</tr>
<tr>
<td>Long Ad Hoc Program</td>
<td>40</td>
</tr>
<tr>
<td>Long Repeated Program</td>
<td>35</td>
</tr>
</tbody>
</table>
The distribution by types of programs for companies reporting types of programs is shown below in Table VI.

TABLE VI

DISTRIBUTION OF COMPANIES ACCORDING TO TYPES OF OUTSIDE PROGRAMS REPORTED

<table>
<thead>
<tr>
<th>Types of Programs</th>
<th>Companies Reporting Outside Programs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
</tr>
<tr>
<td>Short and Long Ad Hoc and Repeated Programs</td>
<td>25</td>
</tr>
<tr>
<td>Short Ad Hoc Program Only</td>
<td>2</td>
</tr>
<tr>
<td>Short and Long Ad Hoc Programs</td>
<td>8</td>
</tr>
<tr>
<td>Short and Long Ad Hoc and Short Repeated Programs</td>
<td>2</td>
</tr>
<tr>
<td>Short and Long Repeated Programs</td>
<td>2</td>
</tr>
<tr>
<td>Long Ad Hoc Program Only</td>
<td>1</td>
</tr>
<tr>
<td>Long Ad Hoc and Long Repeated Programs</td>
<td>5</td>
</tr>
<tr>
<td>Long Repeated Program Only</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
</tr>
</tbody>
</table>

Of the companies which described the types of outside programs in which they participate, thirty-seven said that they made some effort to evaluate outside programs. Nine companies said that they made no effort to evaluate outside programs participated in by them. Four companies of the fifty did not respond to this question. Methods used to evaluate outside programs were
described by thirty-six of the companies.

**Inside Management Development Programs.** Fifty-four companies, representing 94.7% of all companies with formal management development programs, reported that they conduct inside management development programs. Of this number, fifty (92.6%) reported the types of inside programs conducted by them. Table VII below summarizes this information.

**TABLE VII**

**TYPES OF INSIDE PROGRAMS REPORTED BY FIFTY COMPANIES**

<table>
<thead>
<tr>
<th>Types of Programs</th>
<th>Companies Reporting Inside Programs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
</tr>
<tr>
<td>Short <strong>Ad Hoc</strong> Program</td>
<td>37</td>
</tr>
<tr>
<td>Short Repeated Program</td>
<td>40</td>
</tr>
<tr>
<td>Long <strong>Ad Hoc</strong> Program</td>
<td>32</td>
</tr>
<tr>
<td>Long Repeated Program</td>
<td>44</td>
</tr>
</tbody>
</table>

The distribution, by types of programs, for companies reporting types of inside programs is shown as follows in Table VIII.
### TABLE VIII

**DISTRIBUTION OF FIFTY COMPANIES ACCORDING TO TYPES OF INSIDE PROGRAMS REPORTED BY EACH**

<table>
<thead>
<tr>
<th>Type</th>
<th>Number</th>
<th>Per Cent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short and Long Ad Hoc and Repeated Programs</td>
<td>26</td>
<td>52</td>
</tr>
<tr>
<td>Short Ad Hoc Programs Only</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Short Ad Hoc and Short Repeated Programs</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Short Ad Hoc and Short and Long Repeated Programs</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>Short and Long Ad Hoc and Long Repeated Programs</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Short and Long Ad Hoc Programs</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Short Repeated Programs Only</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Long Ad Hoc Programs and Short and Long Repeated Programs</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Short and Long Repeated Programs</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Long Ad Hoc Programs Only</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Long Ad Hoc and Long Repeated Programs</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Long Repeated Programs Only</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

Forty of those companies which reported the types of inside management development programs conducted by them said they make some effort to evaluate the programs. Five companies with inside management development programs said
that they do not evaluate them. The remaining five companies did not give responsive answers.

**Characteristics of Program Participation.** Generally speaking, individual companies have a wider variety of types of inside programs than types of outside programs. The aggregate total of types of inside programs reported by fifty companies is 152 while the aggregate total of types of outside programs reported by the same number of companies is 145.

Short and long *ad hoc* and repeated programs apparently are considered almost equally essential to a well-rounded system of development programs; although, as shown in Table IX, there is a slight tendency for companies in the sample to concentrate more on repeated programs, particularly long repeated programs, than on *ad hoc* programs. Fifty-one companies which reported conducting or participating in each type of program are distributed as shown in Table IX.

**TABLE IX**

**COMBINED PROGRAMS: BY FREQUENCY OF REPORTING BY FIFTY-ONE COMPANIES**

<table>
<thead>
<tr>
<th>Type</th>
<th>Number</th>
<th>Per Cent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short <em>Ad Hoc</em> Programs</td>
<td>42</td>
<td>82.4</td>
</tr>
<tr>
<td>Short Repeated Programs</td>
<td>45</td>
<td>88.2</td>
</tr>
<tr>
<td>Long <em>Ad Hoc</em> Programs</td>
<td>43</td>
<td>84.3</td>
</tr>
<tr>
<td>Long Repeated Programs</td>
<td>46</td>
<td>90.2</td>
</tr>
</tbody>
</table>

The pattern of participation changes somewhat when programs are separated by type between inside and outside programs. There we find that companies rely
more heavily upon repeated inside programs than upon repeated outside programs. The relationship is reversed for ad hoc programs. Companies rely somewhat more heavily upon short ad hoc outside programs than they do upon short ad hoc inside programs and much more heavily on long ad hoc outside programs than they do upon long ad hoc inside programs. This condition is shown in detail by Table X below.

**TABLE X**

**PARTICIPATION:**
INSIDE AND OUTSIDE TYPES OF PROGRAMS, BY FREQUENCY

<table>
<thead>
<tr>
<th>Types of Programs</th>
<th>Inside</th>
<th>Outside</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Per Cent</td>
</tr>
<tr>
<td>Short Ad Hoc</td>
<td>37</td>
<td>74</td>
</tr>
<tr>
<td>Short Repeated</td>
<td>40</td>
<td>80</td>
</tr>
<tr>
<td>Long Ad Hoc</td>
<td>32</td>
<td>64</td>
</tr>
<tr>
<td>Long Repeated</td>
<td>44</td>
<td>88</td>
</tr>
</tbody>
</table>

1Taken from Table VII.
2Taken from Table V.

The foregoing, however, is no more than a tendency. When the distribution by individual companies is taken into account, it is found that the majority of companies have both inside and outside programs of the same general types. Those which have a given type of inside program but no corresponding type of outside program, or vice versa, form a minority, albeit a substantial one where certain types of programs are concerned. Table XI gives the distribution of
types of inside and outside programs by individual companies.

TABLE XI

DISTRIBUTION OF COMPANIES
BY CORRESPONDING TYPES OF INSIDE AND OUTSIDE PROGRAMS

<table>
<thead>
<tr>
<th>Types of Programs</th>
<th>Short Ad Hoc</th>
<th>Short Repeated</th>
<th>Long Ad Hoc</th>
<th>Long Repeated</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>Per Cent</td>
<td>No.</td>
<td>Per Cent</td>
<td>No.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Both</td>
<td>33</td>
<td>64.7</td>
<td>26</td>
<td>51.0</td>
</tr>
<tr>
<td>Inside Only</td>
<td>4</td>
<td>8.0</td>
<td>14</td>
<td>27.5</td>
</tr>
<tr>
<td>Outside Only</td>
<td>5</td>
<td>9.8</td>
<td>5</td>
<td>9.8</td>
</tr>
<tr>
<td>Neither</td>
<td>9</td>
<td>17.6</td>
<td>6</td>
<td>11.8</td>
</tr>
</tbody>
</table>

Companies in the sample tended to show more concern about evaluating inside programs than outside programs. Forty-seven companies reported evaluation of inside and outside programs as shown in Table XII.
TABLE XII

RELATIVE EXTENT TO WHICH INSIDE AND OUTSIDE PROGRAMS ARE EVALUATED BY FORTY-SEVEN COMPANIES

<table>
<thead>
<tr>
<th>Companies Participate In</th>
<th>Number</th>
<th>Per Cent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Both Inside and Outside Programs-- Evaluate Both</td>
<td>35</td>
<td>74.5</td>
</tr>
<tr>
<td>Both Inside and Outside Programs-- Evaluate Inside Only</td>
<td>6</td>
<td>12.8</td>
</tr>
<tr>
<td>Both Inside and Outside Programs-- Do not Evaluate</td>
<td>3</td>
<td>6.3</td>
</tr>
<tr>
<td>Inside Programs Only-- Do not Evaluate</td>
<td>1</td>
<td>2.1</td>
</tr>
<tr>
<td>Outside Programs Only-- Do not Evaluate</td>
<td>2</td>
<td>4.3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>47</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

In other words, 91% (forty-one out of forty-five) of the companies with inside programs make some sort of effort to evaluate them while only 76% (thirty-five out of forty-six) of the companies with outside programs make efforts to evaluate them.

When programs are considered by type, it is found that long or repeated programs are somewhat more likely to be evaluated than short or ad hoc programs. Of those companies which participate in short outside programs, 89.3% evaluate short ad hoc programs and 95.7% evaluate short repeated programs. On the other hand, of those companies which participate in long outside programs, 93.1% evaluate long ad hoc programs and 100% evaluate long repeated programs. Of those companies with inside repeated programs, all said that they evaluate
short repeated programs and 94.1% said they evaluate long repeated programs.

Of those companies with inside *ad hoc* programs, 96.6% said that they evaluated short *ad hoc* programs and 95.7% reported that they evaluate long *ad hoc* programs.

Despite the fact that a high proportion of types of outside programs is evaluated, in only one instance (long repeated programs) was a type of outside program evaluated by as many or more companies as its inside counterpart.

Table XIII, which follows immediately, shows the difference in frequency of evaluation by types of programs.

**TABLE XIII**

EVALUATIONS OF TYPES OF INSIDE AND OUTSIDE PROGRAMS: DIFFERENCES IN FREQUENCIES, BY PERCENTAGE

<table>
<thead>
<tr>
<th>Types of Programs</th>
<th>Percentage Difference, Outside From Inside</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short <em>Ad Hoc</em></td>
<td>-7.3</td>
</tr>
<tr>
<td>Short Repeated</td>
<td>-4.3</td>
</tr>
<tr>
<td>Long <em>Ad Hoc</em></td>
<td>-2.6</td>
</tr>
<tr>
<td>Long Repeated</td>
<td>+5.9</td>
</tr>
</tbody>
</table>

By using the foregoing data, a composite picture can be drawn of the activities of the typical company in formal management development programs. The typical company relies heavily upon programs conducted by itself to handle material which must be taught to successive groups of managerial employees. However, when a special or non-recurring type of need arises, the company will place somewhat more reliance upon programs conducted by outside institutions to meet the need than upon its own programs.

The typical company shows quite a bit of concern about evaluating inside
programs and it is more concerned about evaluating repeated inside programs than it is about ad hoc inside programs. On the other hand, while it is not quite so concerned about evaluating outside programs, when it does decide to evaluate them, it concentrates its efforts more upon repeated programs than it does upon ad hoc programs.

Educational Standards in the Management Development Department. Out of the thirty-seven companies which furnished information about the number of persons with advanced degrees holding responsible positions in management development, only four (12%) said that the possession of an advanced degree in a work-related field is a prerequisite to employment in such positions. One of the four companies said that the advanced degree could be in personnel, industrial relations, psychology, or business administration; a second said that the degree could be in personnel, industrial relations, sociology, or psychology; a third said that the degree should be in business administration; the fourth said that the degree could be in personnel, industrial relations, education, psychology, or business administration.

The academic level of persons holding responsible positions in management development varied widely from company to company. In the majority of companies, no one in a responsible position held an advanced degree. However, a substantial minority of companies had filled responsible positions with persons holding advanced degrees. The percentage of such persons in each company of this group varied widely. (Two companies had 5% and three had 100%. The remainder fell somewhere in between.) Table XIV, which follows immediately, shows the distribution of companies according to the per cent of persons with advanced degrees who hold responsible positions in management development.
TABLE XIV

DISTRIBUTION OF THIRTY-SEVEN COMPANIES BY EMPLOYMENT OF PERSONS WITH ADVANCED DEGREES

<table>
<thead>
<tr>
<th>Per Cent of Persons With Advanced Degrees In Work Related Fields</th>
<th>Companies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
</tr>
<tr>
<td>90-100</td>
<td>3</td>
</tr>
<tr>
<td>80-89</td>
<td>2</td>
</tr>
<tr>
<td>70-79</td>
<td>0</td>
</tr>
<tr>
<td>60-69</td>
<td>1</td>
</tr>
<tr>
<td>50-59</td>
<td>2</td>
</tr>
<tr>
<td>40-49</td>
<td>1</td>
</tr>
<tr>
<td>30-39</td>
<td>2</td>
</tr>
<tr>
<td>20-29</td>
<td>2</td>
</tr>
<tr>
<td>10-19</td>
<td>2</td>
</tr>
<tr>
<td>1-10</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>20</td>
</tr>
</tbody>
</table>

Total With Degrees 17 45.9
Total Without Degrees 20 54.1

Recruitment, Educational Preferences. Eleven out of thirty-seven companies said that they had no preference as to academic background for new employees in management development departments. The twenty companies which expressed
a preference, were about evenly divided between those who preferred an academic background related to business and those who looked for people with academic degrees related to management development. Table XV shows the preferences of the twenty companies as expressed in types of business-related and management development-related degrees.

TABLE XV

DISTRIBUTION OF COMPANIES BY PREFERENCE IN ACADEMIC DEGREES

<table>
<thead>
<tr>
<th>Type of Degree</th>
<th>Choice</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st</td>
<td>2nd</td>
<td>3rd</td>
<td>4th</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>No.</td>
<td>%</td>
<td>No.</td>
<td>%</td>
<td>No.</td>
<td>%</td>
<td>No.</td>
</tr>
<tr>
<td>Business Related</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical</td>
<td>4</td>
<td>20.0</td>
<td>1</td>
<td>6.3</td>
<td>1</td>
<td>6.6</td>
<td>1</td>
</tr>
<tr>
<td>Business</td>
<td>4</td>
<td>20.0</td>
<td>7</td>
<td>43.8</td>
<td>5</td>
<td>33.3</td>
<td>1</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>10.0</td>
<td>1</td>
<td>6.3</td>
<td>3</td>
<td>20.0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>50.0</td>
<td>9</td>
<td>56.4</td>
<td>9</td>
<td>59.9</td>
<td>3</td>
</tr>
<tr>
<td>Management Development Related</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industrial Relations</td>
<td>3</td>
<td>15.0</td>
<td>3</td>
<td>18.7</td>
<td>3</td>
<td>20.0</td>
<td>0</td>
</tr>
<tr>
<td>Education</td>
<td>2</td>
<td>10.0</td>
<td>2</td>
<td>12.5</td>
<td>2</td>
<td>13.3</td>
<td>1</td>
</tr>
<tr>
<td>Psychology</td>
<td>5</td>
<td>25.0</td>
<td>2</td>
<td>12.5</td>
<td>1</td>
<td>6.6</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>50.0</td>
<td>7</td>
<td>43.7</td>
<td>6</td>
<td>39.9</td>
<td>2</td>
</tr>
</tbody>
</table>
It is interesting to note that companies which expressed more than one preference were, for the most part, likely to mix indiscriminately business-related and management development-related degrees. Only four companies out of the twenty consistently showed a preference for business-related or management development-related degrees. The remainder named degrees from both areas as their preferences.

There appears to be some tendency towards professionalization in management development. A few companies have established possession of an advanced degree in management development-related areas as a prerequisite for employment on responsible management development positions. In addition, a substantial minority of companies have filled such positions with persons holding advanced degrees in management development-related areas, without, however, making advanced degrees a prerequisite. Nevertheless, the mixed preference for new employees with business-related and management development-related degrees indicates that the progress of management development towards professional status is more drift than planned change.

**Preference in Work Experience.** Companies which expressed their views about the kind of work experience they look for when hiring new employees for positions in management development, tended to favor persons with experience in areas directly related to management development over those with experience in other areas of business. Table XVI shows the distribution of companies according to preferences for given types of work experience.
### TABLE XVI

**DISTRIBUTION OF TWENTY-SIX COMPANIES BY PREFERENCE IN WORK EXPERIENCE**

<table>
<thead>
<tr>
<th>Type of Work Experience</th>
<th>Choice</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>Per Cent</td>
<td>No.</td>
<td>Per Cent</td>
<td>No.</td>
</tr>
<tr>
<td>Business-Related</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management</td>
<td>4</td>
<td>15.5</td>
<td>7</td>
<td>35.0</td>
<td>1</td>
</tr>
<tr>
<td>Communication&lt;sup&gt;1&lt;/sup&gt;</td>
<td>3</td>
<td>11.5</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Technical</td>
<td>3</td>
<td>11.5</td>
<td>2</td>
<td>10.0</td>
<td>3</td>
</tr>
<tr>
<td>Sales</td>
<td>1</td>
<td>3.8</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>11</td>
<td>42.3</td>
<td>9</td>
<td>45.0</td>
<td>6</td>
</tr>
<tr>
<td>Management Development-Related</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industrial Relations</td>
<td>7</td>
<td>26.9</td>
<td>5</td>
<td>25.0</td>
<td>1</td>
</tr>
<tr>
<td>Behavioral Sciences</td>
<td>2</td>
<td>7.8</td>
<td>3</td>
<td>15.0</td>
<td>2</td>
</tr>
<tr>
<td>Teaching</td>
<td>1</td>
<td>3.8</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Employee Training</td>
<td>5</td>
<td>26.9</td>
<td>5</td>
<td>25.0</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>15</td>
<td>65.4</td>
<td>13</td>
<td>65.0</td>
<td>7</td>
</tr>
</tbody>
</table>

<sup>1</sup> "Communication" work is any type of work which predominately involves handling or communicating with small groups.
Work experience in areas related to management development appears to be more highly valued than formal education in the same areas. Not only does experience in the former areas rank as the first choice of the majority of companies, but, as was not the case with formal education, it ranks above business-related experience as second, third, and fourth choices of the majority of companies.

When the tendency to favor management development-related work experience over business-related work experience is coupled with 50-50 division between those favoring management development-related academic degrees and those favoring business-related academic degrees, a rather strong force is generated which operates to professionalize management development. This would not be the case if the selection of new employees were based entirely on educational preferences. The pattern of work preference makes the difference. The reasons for this will become clear when it is recalled that of those companies which list as first choice persons who have had experience in areas related to management development, three out of fifteen look for persons whose experience (teaching and behavioral sciences) requires advanced degrees and is gained from outside the business world. Assuming that persons who have gained experience in management development-related areas by working in the business world are evenly divided between those holding degrees in business-related and management development-related areas, the influx of persons (who generally possess advanced management development-related degrees) from the teaching profession and the behavioral sciences will serve to increase steadily the number of persons in management development who hold advanced degrees in areas related to management development.
Acceptance of Formal Management Development Programs by Top Management. The person filling out the questionnaire was requested to indicate on a seven-point scale that proportion of top management which believes that management development programs contribute more to the company's welfare than they cost. Thirty-nine companies filled out this section. Their answers are summarized below in Table XVII.

**TABLE XVII**

**ACCEPTANCE OF FORMAL MANAGEMENT DEVELOPMENT PROGRAMS IN THIRTY-NINE COMPANIES**

<table>
<thead>
<tr>
<th>Proportion of Top Management</th>
<th>Companies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
</tr>
<tr>
<td>All</td>
<td>3</td>
</tr>
<tr>
<td>Most</td>
<td>26</td>
</tr>
<tr>
<td>More Than One-Half</td>
<td>5</td>
</tr>
<tr>
<td>One-Half</td>
<td>4</td>
</tr>
<tr>
<td>Less Than One-Half</td>
<td>1</td>
</tr>
</tbody>
</table>

1No company gave answers in the "few" or "none" category.

The foregoing indicates that acceptance of formal management development programs by top management is at a satisfactory level in approximately three-fourths of the companies making reply to the question. In the remaining one-fourth of the companies, acceptance is at a more or less unsatisfactory level.

Financial Stability of Formal Management Development. An important indicator of the degree to which management development programs are considered
essential to the company's welfare is the extent to which the program budget is cut relative to other expenditures (such as research, advertising, public relations, maintenance, etc.) when the company finds it necessary to economize. A question designed to determine this was included in the questionnaire. Thirty-nine companies furnished answers to the question. Of these, the majority indicated that the financial stability of management development programs was as high or higher than other phases of company activity. Their answers are summarized in Table XVIII.

TABLE XVIII

FINANCIAL STABILITY OF PROGRAMS IN THIRTY-NINE COMPANIES

<table>
<thead>
<tr>
<th>Amount of Reduction Relative to Other Reductions</th>
<th>Companies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
</tr>
<tr>
<td>No Reduction</td>
<td>4</td>
</tr>
<tr>
<td>Much Less</td>
<td>4</td>
</tr>
<tr>
<td>Slightly Less</td>
<td>2</td>
</tr>
<tr>
<td>Same</td>
<td>24</td>
</tr>
<tr>
<td>Slightly More</td>
<td>3</td>
</tr>
<tr>
<td>Much More</td>
<td>2</td>
</tr>
</tbody>
</table>

There is a definite relationship between the acceptance of formal management development programs by top management and their stability in times of financial retrenchment. Companies which reported that all of top management accepted the programs also reported that in no case were expenditures for programs reduced more than other company expenditures and, in most cases, they
reported that they were reduced less. The companies which reported that one-half or less of top management accepted the programs, also reported experiencing larger reductions in program expenditure than those made in the other areas of company operation.

The relationship between acceptability and financial stability is shown as follows in the next table. The acceptability and financial stability categories have been assigned numerical values with one equaling the least amount of acceptability and stability and six equaling the greatest amount of acceptability and stability.

**TABLE XIX**

RELATIONSHIP BETWEEN ACCEPTABILITY AND FINANCIAL STABILITY IN THIRTY-SEVEN COMPANIES

<table>
<thead>
<tr>
<th>Acceptance Score</th>
<th>Average Stability Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td>1.0</td>
</tr>
<tr>
<td>3.</td>
<td>3.0</td>
</tr>
<tr>
<td>4.</td>
<td>3.2</td>
</tr>
<tr>
<td>5.</td>
<td>3.5</td>
</tr>
<tr>
<td>6.</td>
<td>5.0</td>
</tr>
</tbody>
</table>

When an economy drive is on, the training budget usually is hardest hit in the area of outside programs. In response to a question about the degree to which expenditures for outside programs are reduced relative to inside programs, thirty-six companies answered as follows: Seventeen (47.2%) reduce outlay for outside programs to the same extent as for inside programs; five (13.9%) reported that expenditures for inside programs are reduced to a greater extent than those for outside programs; however, fourteen (38.9%) said that expenditures for outside programs are reduced more than those for inside programs.
Concluding Statement. The great majority of companies who responded to the questionnaire have formal management development programs. Those with such programs tended to become involved in a greater variety of inside programs than outside programs. Most companies rely to a greater extent upon outside organizations to meet special or non-recurring needs than they do upon their own training resources. Conversely, most relied more heavily upon their own training organization to handle repeated programs than they did upon outside organizations.

The majority of companies which gave answers to questions about the evaluation of programs, indicated that they make efforts to evaluate both inside and outside programs. However, outside programs are evaluated with somewhat less frequency than inside programs. Whether or not a given program is repeated seems to be the single most important factor in determining it will be evaluated.

Of those companies which answered questions relating to acceptance of formal management development programs by top management, a majority indicated that most of top management believe that management development programs contribute more to the welfare of the company than they cost.

The financial stability of management development programs depends closely upon their acceptance by top management. In general, it appears to be equal to or better than, that of other phases of company operation. Companies in the sample said, for the most part, that when expenses are cut generally, management development program budgets either are not cut as heavily as other items of expense or are cut the same. It is interesting to note, however, that between inside and outside programs, the latter usually bear more than their share of the reduction when expenses are cut.
CHAPTER VIII

PROCEDURAL EFFECTIVENESS

In this chapter we will explore the procedural effectiveness of evaluation plans found in the sample. The reader will recall that, in order to be completely effective from the standpoint of procedure, an evaluation plan must: (1) Evaluate outside programs, where present; (2) evaluate inside programs, where present; (3) evaluate teaching methods when they are within the control of the company; (4) use evaluational data as the basis for decisions about whether to continue, modify, or discontinue programs and teaching methods. Keeping this in mind we will now determine the procedural effectiveness of evaluation plans as they are used in connection with inside programs, outside programs, and teaching methods.

Outside Programs. Thirty-five companies reported which types of outside programs they evaluate. The following, Table XX, summarizes this information.
<table>
<thead>
<tr>
<th>Type of Program</th>
<th>Companies Have Program</th>
<th>Companies Evaluate Program</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>Per Cent</td>
</tr>
<tr>
<td>Short Ad Hoc Programs</td>
<td>27</td>
<td>77.1</td>
</tr>
<tr>
<td>Short Repeated Programs</td>
<td>23</td>
<td>65.7</td>
</tr>
<tr>
<td>Long Ad Hoc Programs</td>
<td>29</td>
<td>82.9</td>
</tr>
<tr>
<td>Long Repeated Programs</td>
<td>28</td>
<td>80.0</td>
</tr>
</tbody>
</table>

After a score of ten is assigned to each category of programs and deductions are made from that score in accordance with negative scores (shown in Table I found at page 62) which have been adjusted to reflect the different number of companies in each of the four categories, we arrive at scores for procedural effectiveness in evaluating outside programs as shown in Table XXI.
TABLE XXI
PROCEDURAL SCORES: EVALUATION
OF OUTSIDE PROGRAMS BY THIRTY-FIVE COMPANIES

<table>
<thead>
<tr>
<th>Type of Program</th>
<th>Negative Score</th>
<th>Average Deduction Per Program</th>
<th>Per Cent of Mean of all Programs Reported</th>
<th>Adjusted Deductions¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short Ad Hoc</td>
<td>1</td>
<td>.111</td>
<td>101.0</td>
<td>.112</td>
</tr>
<tr>
<td>Short Repeated</td>
<td>3</td>
<td>.13</td>
<td>85.8</td>
<td>.112</td>
</tr>
<tr>
<td>Long Ad Hoc</td>
<td>2</td>
<td>.138</td>
<td>108.0</td>
<td>.149</td>
</tr>
<tr>
<td>Long Repeated</td>
<td>4</td>
<td>0</td>
<td>80.0</td>
<td>0</td>
</tr>
<tr>
<td>Total Deduction</td>
<td></td>
<td></td>
<td></td>
<td>.373</td>
</tr>
<tr>
<td>Net Score (10 - .373)</td>
<td></td>
<td></td>
<td>9.627</td>
<td></td>
</tr>
</tbody>
</table>

¹Adjusted deduction is calculated by adding negative scores in a given category, dividing them by the number of companies having programs in that category and multiplying the result by the percentage that the number of companies in that category represented as compared with the average number of companies in all categories.

The companies in the sample, as shown by Table XXI, do an excellent job of including all types of outside programs in their evaluational plans. However, this is not the case when it comes to using evaluational data as the basis for making decisions about modifying or abandoning outside repeated programs. Twenty-two companies reported evaluating short repeated outside programs. Of this number, only twelve (54.5%) use the evaluational data thus obtained as the primary basis for making major administrative decisions about the programs. Twenty-eight companies reported evaluating long repeated programs.
outside programs. Only eighteen (64.3%) said that evaluation data from this source is used as the primary basis for making major administrative decisions about long repeated outside programs.

After a score of ten is assigned to each category of repeated outside programs and deductions are made from that score in accordance with negative scores (shown in Table II found at page 64) which have been adjusted to reflect the different number of companies in each category of repeated outside programs, we arrive at the net scores for using evaluation data shown in Table XXII.

**TABLE XXII**

USE OF EVALUATION DATA FROM OUTSIDE PROGRAMS AS REPORTED BY THIRTY-FIVE COMPANIES

<table>
<thead>
<tr>
<th>Type of Program</th>
<th>Negative Score</th>
<th>Average Deduction Per Program</th>
<th>% of Mean, All Repeated Programs Reported</th>
<th>Adjusted Deduction�</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short Repeated</td>
<td>4.3</td>
<td>1.95</td>
<td>88.0</td>
<td>1.72</td>
</tr>
<tr>
<td>Long Repeated</td>
<td>5.7</td>
<td>2.04</td>
<td>112.0</td>
<td>2.45</td>
</tr>
<tr>
<td>Total Deduction</td>
<td></td>
<td></td>
<td></td>
<td>4.17</td>
</tr>
<tr>
<td>Net Score (10 - 4.17)</td>
<td></td>
<td></td>
<td></td>
<td>5.83</td>
</tr>
</tbody>
</table>

�Adjusted deduction is calculated by adding negative scores in a given category, dividing them by the number of companies having programs in that category and multiplying the result by the percentage that the number of companies in that category represented as compared with the average number of companies in all repeated program categories.

Evaluation of Inside Programs. Thirty-nine companies reported the types of inside programs they evaluate. The following table summarizes this information.
**TABLE XXIII**

TYPES OF INSIDE PROGRAMS EVALUATED BY THIRTY-NINE COMPANIES

<table>
<thead>
<tr>
<th>Type of Program</th>
<th>Companies Have Program</th>
<th>Companies Evaluate Program</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>Per Cent</td>
</tr>
<tr>
<td>Short Ad Hoc Programs</td>
<td>29</td>
<td>74.4</td>
</tr>
<tr>
<td>Short Repeated Programs</td>
<td>32</td>
<td>82.1</td>
</tr>
<tr>
<td>Long Ad Hoc Programs</td>
<td>23</td>
<td>59.0</td>
</tr>
<tr>
<td>Long Repeated Programs</td>
<td>34</td>
<td>87.2</td>
</tr>
</tbody>
</table>

By following the same procedure used to construct Table XXI, we arrive at the scores for coverage of inside programs summarized by Table XXIV.
TABLE XXIV
PROCEDURAL SCORES: EVALUATION OF INSIDE PROGRAMS BY THIRTY-NINE COMPANIES

<table>
<thead>
<tr>
<th>Type of Program</th>
<th>Negative Score</th>
<th>Average Deduction Per Program</th>
<th>Per Cent of Mean Of All Programs Reported</th>
<th>Adjusted Deduction ¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short Ad Hoc</td>
<td>1</td>
<td>.0345</td>
<td>98.3</td>
<td>.034</td>
</tr>
<tr>
<td>Short Repeated</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Long Ad Hoc</td>
<td>2</td>
<td>.0869</td>
<td>78.0</td>
<td>.068</td>
</tr>
<tr>
<td>Long Repeated</td>
<td>4</td>
<td>.235</td>
<td>115.2</td>
<td>.271</td>
</tr>
</tbody>
</table>

Total Deduction  |
Net Score (10 - .373)  | 9.627

¹Adjusted deduction was calculated in the manner described in footnote 1 to Table XXI.

Although the companies did a thorough job of covering all types of inside programs in their evaluational plans, they did not do so well in putting evaluation data to use in making decisions about modifying or abandoning inside repeated programs. Thirty-two companies said that they evaluate short repeated inside programs. Of this number, only twenty (62.5%) said they use the evaluation data thus obtained as the primary basis for making important administrative decisions about the programs. Thirty-two companies reported evaluating long repeated inside programs. Only twenty-one (65.6%) said that evaluation data from this source was used as the primary basis for making important administrative decisions about long repeated inside programs.
Following the same procedure used to construct Table XXII, we arrive at the procedural scores for use of evaluation data of repeated inside programs shown below in Table XXV.

**TABLE XXV**

**USE OF EVALUATION DATA FROM INSIDE PROGRAMS AS REPORTED BY THIRTY-NINE COMPANIES**

<table>
<thead>
<tr>
<th>Type of Program</th>
<th>Negative Score</th>
<th>Average Deduction Per Program</th>
<th>Per Cent of Mean, All Repeated Programs Reported</th>
<th>Adjusted Deduction^1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short Repeated</td>
<td>4.3</td>
<td>1.61</td>
<td>100.0</td>
<td>1.61</td>
</tr>
<tr>
<td>Long Repeated</td>
<td>5.7</td>
<td>1.96</td>
<td>100.0</td>
<td>1.96</td>
</tr>
<tr>
<td>Total Deduction</td>
<td></td>
<td></td>
<td></td>
<td>3.57</td>
</tr>
<tr>
<td>Net Score (10 - 3.57)</td>
<td></td>
<td></td>
<td></td>
<td>6.43</td>
</tr>
</tbody>
</table>

^1 Adjusted deduction was calculated in the manner described in footnote 1 to Table XXII.

Comparisons of Procedural Effectiveness, Inside and Outside Programs. Deductions made in the first two sections of this chapter for failure to evaluate specified types of outside and inside programs are summarized below.
TABLE XXVI

COMPARISONS BETWEEN DEDUCTIONS MADE FOR FAILURE TO EVALUATE INSIDE AND OUTSIDE PROGRAMS

<table>
<thead>
<tr>
<th>Type of Program</th>
<th>Adjusted Deduction Outside Program</th>
<th>Adjusted Deduction Inside Program</th>
<th>Difference, Outside Minus Inside Deductions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short Ad Hoc</td>
<td>.112</td>
<td>.034</td>
<td>.078</td>
</tr>
<tr>
<td>Short Repeated</td>
<td>.112</td>
<td>0</td>
<td>.112</td>
</tr>
<tr>
<td>Long Ad Hoc</td>
<td>.149</td>
<td>.068</td>
<td>.081</td>
</tr>
<tr>
<td>Long Repeated</td>
<td>0</td>
<td>.271</td>
<td>-.271</td>
</tr>
<tr>
<td>Total</td>
<td>.373</td>
<td>.373</td>
<td></td>
</tr>
<tr>
<td>Net Score (10 - .373)</td>
<td>9.627</td>
<td>9.627</td>
<td></td>
</tr>
</tbody>
</table>

1. Taken from Table XXI.
2. Taken from Table XXIV.

It is evident that companies have a greater tendency to evaluate short Ad Hoc, short repeated and long Ad Hoc inside programs than they do to evaluate corresponding types of outside programs. However, the reverse is true of long repeated programs. The upshot is that procedural scores for including inside and outside programs in evaluational plans were the same.

We are confronted with a different situation, however, when we compare the number of companies which reported putting to use evaluation data from outside programs with those reporting on the same subject for inside...
In this case, evalutional data from inside programs are used as the basis for important administrative decisions more often than are those from outside programs. Comparisons between procedural scores in this area for inside and outside programs are shown in the following table identified as Table XXVII.

**TABLE XXVII**

**COMPARISONS BETWEEN DEDUCTIONS MADE FOR FAILURE TO PUT TO USE EVALUATIONAL DATA FROM INSIDE AND OUTSIDE PROGRAMS**

<table>
<thead>
<tr>
<th>Type of Program</th>
<th>Adjusted Deduction Outside Program</th>
<th>Adjusted Deduction Inside Program</th>
<th>Difference, Outside Minus Inside Deductions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short Repeated</td>
<td>1.72</td>
<td>1.61</td>
<td>.11</td>
</tr>
<tr>
<td>Long Repeated</td>
<td>2.45</td>
<td>1.96</td>
<td>.49</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>4.17</strong></td>
<td><strong>3.57</strong></td>
<td><strong>.60</strong></td>
</tr>
<tr>
<td><strong>Net Scores</strong></td>
<td><strong>5.83</strong></td>
<td><strong>6.43</strong></td>
<td></td>
</tr>
<tr>
<td>(10 - 4.17 and 10 - 3.57)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Taken from Table XXII.

2 Taken from Table XXV.

The tendency for companies to use the findings of evaluations made of inside programs as the basis for administrative decisions more often than they do evaluations of outside programs gives an edge to the procedural effectiveness scores for inside programs over those for outside programs. The total average net score for inside programs of all companies is 16.057
out of a maximum possible score of twenty as compared with 14.457 for outside programs.

**Evaluation of Teaching Methods, Procedural.** Forty-three companies furnished information relative to the evaluation of teaching methods. Of this number, thirty-two (74.4%) said that they evaluate teaching methods and eleven (25.5%) said that they do not. Of those companies who said that they evaluate teaching methods, twenty-two (68.8%) said that they used the results of their evaluations as the basis for administrative decisions relating to changing, continuing, or modifying teaching methods. The remaining ten companies (31.2%) said that they did not use evaluation data for this purpose.

By assigning ten points for evaluating teaching methods and ten points for putting the evaluation data to use and making deductions (weighted to reflect the number of companies in each category) for failure to make or use evaluations, we find that the net procedural scores for thirty-two companies in this area are: Evaluation of teaching methods, 7.44, and use of evaluation data, 6.87.

**Procedural Scores, Total.** By considering and weighing answers of companies to specific categories of questions (without regard to the fact that certain companies may not have supplied answers to all relevant categories), we find that the total average procedural score based on answers to all relevant categories of questions is 44.8 points out of a possible sixty points. However, the average procedural score for thirty-nine companies, each of which supplied answers to all relevant categories of questions, is
42.1 points out of a possible sixty points. The distribution of these companies, by procedural scores, is shown in Table XXVIII which follows immediately.

### TABLE XXVIII

DISTRIBUTION OF PROCEDURAL SCORES FOR THIRTY-NINE COMPANIES

<table>
<thead>
<tr>
<th>Procedural Score Ranges</th>
<th>Number</th>
<th>Per Cent</th>
</tr>
</thead>
<tbody>
<tr>
<td>56 - 60</td>
<td>13</td>
<td>33.3</td>
</tr>
<tr>
<td>51 - 55</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>46 - 50</td>
<td>4</td>
<td>10.3</td>
</tr>
<tr>
<td>41 - 45</td>
<td>2</td>
<td>4.1</td>
</tr>
<tr>
<td>36 - 40</td>
<td>6</td>
<td>15.4</td>
</tr>
<tr>
<td>31 - 35</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>26 - 30</td>
<td>7</td>
<td>17.9</td>
</tr>
<tr>
<td>21 - 25</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16 - 20</td>
<td>6</td>
<td>15.4</td>
</tr>
<tr>
<td>11 - 15</td>
<td>1</td>
<td>2.6</td>
</tr>
<tr>
<td>6 - 10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1 - 5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>39</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

1 These scores were computed for each individual company in accordance with the equations found on page 60.
As shown by the table of distribution, a relative large proportion of companies scored in the top range. The remaining companies were widely and fairly evenly distributed throughout the various score ranges.

Concluding Statement. It appears that once a given company decides to evaluate formal management development programs, it includes in its evaluation plan almost all types of inside and outside programs regardless of length or nature (repeated or ad hoc). To a lesser, but still substantial extent, it will include teaching methods in the plan.

There is not as great a tendency to use evaluation results as the basis for making administrative decisions relative to programs and teaching methods. Of those companies which reported evaluating both short and long repeated outside programs, only 54.4% and 64.3%, respectively, said that they use evaluation data thus obtained as the primary basis for such decisions. Of those companies which reported evaluating both short and long repeated inside programs, a somewhat greater proportion, 62.5% and 65.6%, respectively, said that evaluation data are used as the primary basis for administrative decisions.

Answers given by companies about the use of data from evaluations made of teaching methods display a similar pattern. Of those companies which evaluate teaching methods, only 68.8% said that evaluation data are used as the primary basis for administrative decisions about particular teaching methods.

The average procedural score for companies which furnished answers to all questions bearing upon procedure was rather low—42.8 points out of a possible 60 points or, to put it another way, 71.3% of the possible maximum score. It is noteworthy, however, that one-third of the companies earned
procedural scores ranging from fifty-six to sixty, the maximum. This sug-
gests that there may be a single factor or a group of closely related
factors at work to produce a rather well-marked division between the
evaluational effectiveness of certain groups of companies. Later in the
thesis, this possibility will be discussed more fully.
The technical effectiveness of evaluation techniques will be discussed in this chapter. The reader will recall that a system was established in Chapter VI by which techniques used to evaluate programs are rated according to their effectiveness in measuring attainment of: (1) Acceptance of the program by the participants; (2) transmission of information to the participants; and (3) on-the-job improvements resulting from the program. Techniques of evaluating teaching methods were rated according to their capacity to determine the effect of teaching methods upon learning by the participant.

Using data taken from ratings, the effectiveness of evaluation techniques will be discussed; first, as they relate to outside programs; secondly, as they relate to inside programs; third, as they relate to both inside and outside programs; and fourth, as they relate to teaching methods.

**Outside Programs.** Thirty-seven companies reported the types of techniques they use to evaluate outside programs participated in by them.¹ The following table (Table XXIX) gives the distribution of answers according to techniques.

---

¹The discrepancy between the number of companies which reported types of outside programs evaluated and the number of companies which reported techniques used to evaluate outside programs is accounted for by the fact that one company with outside programs failed to specify which types it evaluates although it reported the techniques used to evaluate them.
TABLE XXIX

OUTSIDE PROGRAM EVALUATION TECHNIQUES REPORTED BY THIRTY-SEVEN COMPANIES.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Companies</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Testing Participants Before Commencing and After Completing Program</td>
<td>8</td>
<td>22.2</td>
</tr>
<tr>
<td>(No Control Group)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testing Participants Before Commencing and After Completing Program</td>
<td>2</td>
<td>5.6</td>
</tr>
<tr>
<td>(Control Group)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measurement of Job Performance Before and After Attending Program</td>
<td>11</td>
<td>30.5</td>
</tr>
<tr>
<td>(No Control Group)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rating of Program by Participants</td>
<td>37</td>
<td>100.0</td>
</tr>
<tr>
<td>Rating by Superiors of Participants' Improvement in Job Performance</td>
<td>31</td>
<td>83.8</td>
</tr>
<tr>
<td>After Attending Program (No Control Group)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rating by Peers of Participants' Improvement in Job Performance</td>
<td>6</td>
<td>16.2</td>
</tr>
<tr>
<td>After Attending Program (No Control Group)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Two facts become immediately evident. First, the average company employs not one, but several techniques (2.5 to be exact) to evaluate outside programs. Secondly, the techniques most favored are those which are farthest removed from the controlled experiment type. The latter characteristic will become more evident when the techniques are rated according to their ability to measure progress toward the three general program criteria.

Using the technical rating system discussed at pages 67 and 68 of Chapter VI, it was found that the above companies averaged 2.9 out of a possible score
of three for techniques used to measure the "acceptability" criterion; .7 out of a possible score of three for techniques used to measure the "transmission of information" criterion; and .7 out of a possible score of three for techniques used to measure the "improvement in job performance" criterion. The distribution of the scores are shown in Table XXX.

**TABLE XXX**

**DISTRIBUTION BY CRITERION OF TECHNICAL SCORES ASSIGNED TO EVALUATION METHODS FOR OUTSIDE PROGRAMS REPORTED BY THIRTY-SEVEN COMPANIES**

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Number and Per Cent of Companies Earning Specified Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0  %</td>
</tr>
<tr>
<td>Acceptability</td>
<td>1</td>
</tr>
<tr>
<td>Transmission of Information</td>
<td>28</td>
</tr>
<tr>
<td>Improvement in Job Performance</td>
<td>5</td>
</tr>
</tbody>
</table>

Table XXX indicates that nearly all of the companies studied make efforts to determine whether outside programs are acceptable to the participants. Moreover, a good proportion make an effort to see to what extent the programs have resulted in improved job performance. However, rather few make serious efforts to find out how much information is transmitted to the participants.

The average score for thirty-seven companies on methods reported by them as being used to evaluate outside programs is 4.3 out of a possible nine
points. The distribution of the scores by numbers of companies is shown next in Table XXXI.

**TABLE XXXI**

**DISTRIBUTION OF TECHNICAL SCORES FOR OUTSIDE PROGRAM EVALUATION METHODS REPORTED BY THIRTY-SEVEN COMPANIES**

<table>
<thead>
<tr>
<th>Score</th>
<th>Companies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
</tr>
<tr>
<td>9.0</td>
<td>0</td>
</tr>
<tr>
<td>8.5</td>
<td>0</td>
</tr>
<tr>
<td>8.0</td>
<td>2</td>
</tr>
<tr>
<td>7.5</td>
<td>0</td>
</tr>
<tr>
<td>7.0</td>
<td>7</td>
</tr>
<tr>
<td>6.5</td>
<td>1</td>
</tr>
<tr>
<td>6.0-4.5</td>
<td>0</td>
</tr>
<tr>
<td>4.0</td>
<td>5</td>
</tr>
<tr>
<td>3.5</td>
<td>18</td>
</tr>
<tr>
<td>3.0</td>
<td>4</td>
</tr>
<tr>
<td>2.5-.5</td>
<td>0</td>
</tr>
</tbody>
</table>

Here we find that scores tend to cluster at two distinct points above and below the mean. The upper cluster is somewhat smaller than the lower and the two are separated by a considerable distance.

**Inside Programs.** Forty-one companies described the techniques used by them to evaluate inside programs. The following table (Table XXXII) gives
the distribution of answers according to evaluation techniques mentioned.

TABLE XXXII

INSIDE PROGRAM EVALUATION TECHNIQUES
REPORTED BY FORTY-ONE COMPANIES

<table>
<thead>
<tr>
<th>Technique</th>
<th>Companies</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Per Cent</td>
</tr>
<tr>
<td>Testing Participants Before Commencing and After Completing Program (No Control Group)</td>
<td>15</td>
<td>36.6</td>
</tr>
<tr>
<td>Testing Participants Before Commencing and After Completing Program (Control Group)</td>
<td>3</td>
<td>7.4</td>
</tr>
<tr>
<td>Measurement of Job Performance Before and After Attending Program (No Control Group)</td>
<td>16</td>
<td>39.0</td>
</tr>
<tr>
<td>Measurement of Job Performance Before and After Attending Program (Control Group)</td>
<td>2</td>
<td>4.9</td>
</tr>
<tr>
<td>Rating of Program by Participants</td>
<td>38</td>
<td>92.7</td>
</tr>
<tr>
<td>Rating by Superiors of Participants' Improvement in Job Performance After Attending Program (No Control Group)</td>
<td>38</td>
<td>92.7</td>
</tr>
<tr>
<td>Rating by Peers of Participants' Improvement in Job Performance After Attending Program (No Control Group)</td>
<td>7</td>
<td>17.0</td>
</tr>
</tbody>
</table>

As shown by Table XXXII, the average company uses 2.9 separate techniques in connection with its evaluation of inside programs (.4 more than are used in connection with outside programs). Although the majority of methods listed by companies as being used to evaluate outside programs are of the informal or non-experimental type, there appears to be a greater tendency to use techniques of the controlled experiment type than was present with outside
programs.

After inside program evaluation techniques were graded in accordance with the technical rating system (see page 67), it was found that companies averaged 2.8 points out of a possible score of three for techniques used to measure the "acceptability" criterion; 1.1 points out of a possible score of three for techniques used to measure the "transmission of information" criterion; and .9 point out of a possible score of three for techniques used to measure the "improvement in job performance" criterion. The distribution of these scores is shown in Table XXXIII.

**TABLE XXXIII**

**DISTRIBUTION BY CRITERION, OF TECHNICAL SCORES FOR EVALUATION METHODS OF OUTSIDE PROGRAMS REPORTED BY FORTY-ONE COMPANIES**

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Number and Per Cent of Companies Earning Specified Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 %</td>
</tr>
<tr>
<td>Acceptability</td>
<td>3</td>
</tr>
<tr>
<td>Transmission of Information</td>
<td>26</td>
</tr>
<tr>
<td>Improvement in Job Performance</td>
<td>2</td>
</tr>
</tbody>
</table>

Data in the table indicate that nearly all companies made efforts to determine (1) whether inside programs are acceptable to the participants and (2) what contribution inside programs made to improvement in job performance.
A substantial minority also made attempts to find out how much information was transmitted to the participants by the programs.

The average score of forty-one companies for methods reported by them as being used to evaluate inside programs is 4.8 points out of a possible nine points. The distribution of the scores by numbers of companies is shown next in Table XXXIV.

**TABLE XXXIV**

**DISTRIBUTION OF TECHNICAL SCORES FOR INSIDE PROGRAM EVALUATION METHODS REPORTED BY FORTY-ONE COMPANIES**

<table>
<thead>
<tr>
<th>Score</th>
<th>Companies</th>
<th>Number</th>
<th>Per Cent</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.0</td>
<td>3</td>
<td>7.3</td>
<td></td>
</tr>
<tr>
<td>8.5</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>8.0</td>
<td>1</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>7.5</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>7.0</td>
<td>7</td>
<td>17.1</td>
<td></td>
</tr>
<tr>
<td>6.5</td>
<td>4</td>
<td>9.7</td>
<td></td>
</tr>
<tr>
<td>6.0</td>
<td>1</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>5.5-4.5</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4.0</td>
<td>6</td>
<td>14.6</td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>17</td>
<td>41.4</td>
<td></td>
</tr>
<tr>
<td>3.0</td>
<td>1</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>2.5-1.5</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>1</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
Again we find that scores tend to cluster at rather widely separated points above and below the mean.

Comparisons of Technical Effectiveness, Outside and Inside Programs. Evaluation techniques, by type, reported by companies in the sample and the percentage of companies reporting the use of each technique to evaluate inside and outside programs are shown in Table XXXV.
### TABLE XXXV

**PERCENTAGE OF COMPANIES USING SPECIFIED EVALUATION TECHNIQUES IN CONJUNCTION WITH OUTSIDE AND INSIDE PROGRAMS**

<table>
<thead>
<tr>
<th>Technique</th>
<th>Per Cent Outside Programs</th>
<th>Per Cent Inside Programs</th>
<th>Difference, Inside Minus Outside</th>
</tr>
</thead>
<tbody>
<tr>
<td>Testing Participants Before Commencing and After Completing Program (No Control Group)</td>
<td>22.2</td>
<td>36.6</td>
<td>14.6</td>
</tr>
<tr>
<td>Testing Participants Before Commencing and After Completing Program (Control Group)</td>
<td>5.6</td>
<td>7.4</td>
<td>1.8</td>
</tr>
<tr>
<td>Measurement of Job Performance Before and After Attending Program (No Control Group)</td>
<td>30.5</td>
<td>39.0</td>
<td>8.5</td>
</tr>
<tr>
<td>Measurement of Job Performance Before and After Attending Program (Control Group)</td>
<td>0</td>
<td>4.9</td>
<td>4.9</td>
</tr>
<tr>
<td>Rating of Program by Participants</td>
<td>100.0</td>
<td>92.7</td>
<td>-8.3</td>
</tr>
<tr>
<td>Rating by Superiors of Participants' Improvement in Job Performance After Attending Program (No Control Group)</td>
<td>86.1</td>
<td>92.7</td>
<td>6.6</td>
</tr>
<tr>
<td>Rating by Peers of Participants' Improvement in Job Performance After Attending Program (No Control Group)</td>
<td>16.2</td>
<td>17.0</td>
<td>1.8</td>
</tr>
</tbody>
</table>

1 Data taken from Table XXIX found at page 104.

2 Data taken from Table XXXII found at page 107.

The data in Table XXXV indicate two things: First, that companies use fewer techniques to evaluate outside programs than they use to evaluate inside programs and, secondly, that they use more techniques of the controlled
experiment type when evaluating inside programs than when evaluating outside programs.

In connection with the first point, it was found that while the average company uses 2.5 types of methods to evaluate outside programs, it uses 2.9 types to evaluate inside programs. In connection with the second point, it was found that while 12.3% of the companies reported using controlled experiment type methods to evaluate inside programs, only 5.6% use the same methods to evaluate outside programs. Moreover, the tendency to use less exacting methods to evaluate outside programs holds true for individual companies which reported types of methods used to evaluate both inside and outside programs as well as for all companies which reported methods used to evaluate outside programs. This is portrayed by Table XXXVI.
TABLE XXXVI

EVALUATION TECHNIQUES REPORTED BY THIRTY-FIVE COMPANIES WHICH EVALUATE BOTH INSIDE AND OUTSIDE PROGRAMS

<table>
<thead>
<tr>
<th>Technique</th>
<th>Companies Use Same Technique for Both Types of Programs</th>
<th>Companies Use Technique for Inside Program Only</th>
<th>Companies Use Technique for Outside Program Only</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>%</td>
<td>No.</td>
</tr>
<tr>
<td>Testing Participants, (No Control Group)</td>
<td>7</td>
<td>20.0</td>
<td>6</td>
</tr>
<tr>
<td>Testing Participants, (Control Group)</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Measurement of Job Performance (No Control Group)</td>
<td>7</td>
<td>20.0</td>
<td>4</td>
</tr>
<tr>
<td>Measurement of Job Performance (Control Group)</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Rating by Participants</td>
<td>33</td>
<td>94.3</td>
<td>0</td>
</tr>
<tr>
<td>Rating by Superiors</td>
<td>30</td>
<td>85.7</td>
<td>3</td>
</tr>
<tr>
<td>Rating by Peers</td>
<td>5</td>
<td>14.3</td>
<td>0</td>
</tr>
</tbody>
</table>

When methods used by companies to evaluate inside and outside programs were compared, it was found that latter methods were somewhat more effective in measuring the "acceptability" criterion, but considerably less effective in measuring "transmission of information" and "improvement in job performance" criteria. This information is summarized in Table XXXVII.
TABLE XXXVII

AVERAGE TECHNICAL SCORES, BY CRITERION, FOR TECHNIQUES USED TO EVALUATE INSIDE AND OUTSIDE PROGRAMS

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Average Score, Techniques, Inside Programs (N = 41)</th>
<th>Average Score, Techniques, Outside Programs (N = 37)</th>
<th>Weighted Average, All Techniques (N = 78)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceptability</td>
<td>2.8</td>
<td>2.9</td>
<td>2.8</td>
</tr>
<tr>
<td>Transmission of Information</td>
<td>1.1</td>
<td>.7</td>
<td>.9</td>
</tr>
<tr>
<td>Improvement Job Performance</td>
<td>.9</td>
<td>.7</td>
<td>.8</td>
</tr>
</tbody>
</table>

When the distribution of scores of individual companies for methods used to evaluate inside and outside programs are compared, it is found that while both cluster at roughly the same levels, a greater proportion of scores for methods used to evaluate inside programs are at the higher level than that of scores for inside program evaluation methods. This is shown in Table XXXVIII.
TABLE XXXVIII

DISTRIBUTION OF TECHNICAL SCORES FOR
OUTSIDE AND INSIDE PROGRAM EVALUATION METHODS

<table>
<thead>
<tr>
<th>Score</th>
<th>Per Cent Outside Individual Scores(^1)</th>
<th>Per Cent Inside Individual Scores(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.0</td>
<td>0</td>
<td>7.3</td>
</tr>
<tr>
<td>8.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8.0</td>
<td>5.4</td>
<td>2.4</td>
</tr>
<tr>
<td>7.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7.0</td>
<td>18.9</td>
<td>17.1</td>
</tr>
<tr>
<td>6.5</td>
<td>2.7</td>
<td>9.7</td>
</tr>
<tr>
<td>6.0</td>
<td>0</td>
<td>2.4</td>
</tr>
<tr>
<td>5.5-4.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4.0</td>
<td>13.5</td>
<td>14.6</td>
</tr>
<tr>
<td>3.5</td>
<td>48.6</td>
<td>41.4</td>
</tr>
<tr>
<td>3.0</td>
<td>10.8</td>
<td>2.4</td>
</tr>
<tr>
<td>2.5-1.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.0</td>
<td>0</td>
<td>2.4</td>
</tr>
<tr>
<td>.5</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\(^1\) Data taken from Table XXXI found at page 106.

\(^2\) Data taken from Table XXXIV found at page 109.

The average scores for methods used to evaluate outside and inside programs, respectively, are 4.3 points and 4.7 points. The average score for outside program evaluation methods is 47.8% of the maximum possible score.
while that for inside program evaluation techniques is 52.2% of the maximum possible score. The weighted average of scores for all techniques used to evaluate both inside and outside programs is 4.5 -- 50% of the maximum possible score.

**Evaluation of Teaching Methods.** Most companies tend to use more than one technique to evaluate teaching methods (the average number of techniques used is 1.5). The majority of companies which described the techniques used by them to evaluate teaching methods, rely primarily upon ratings of methods by participants. Others simply use test results, without control groups, as indicators of the effectiveness of a given method. A somewhat smaller group makes comparisons on the basis of test results between two or more teaching methods. A very small number reported using ratings made by instructors or outside observers to evaluate teaching methods. Table XXXIX gives complete information about the use of various evaluation techniques by the companies in the sample.

**TABLE XXXIX**

**TYPES OF TEACHING METHOD EVALUATION TECHNIQUES REPORTED BY THIRTY-TWO COMPANIES**

<table>
<thead>
<tr>
<th>Technique</th>
<th>Companies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Testing (No Comparisons)</td>
<td>9</td>
</tr>
<tr>
<td>Testing (Comparisons)</td>
<td>5</td>
</tr>
<tr>
<td>Rating by Participants (No Comparisons)</td>
<td>31</td>
</tr>
<tr>
<td>Rating by Instructor (No Comparisons)</td>
<td>1</td>
</tr>
<tr>
<td>Rating by Outside Observers (Comparisons)</td>
<td>1</td>
</tr>
</tbody>
</table>
Techniques used to evaluate teaching methods were rated in accordance with the system described at page 66, Chapter VI. The following table (XL) gives the distribution of rating scores for teaching method evaluation techniques.

### TABLE XL

**DISTRIBUTION OF SCORES FOR TEACHING METHOD EVALUATION TECHNIQUES OF THIRTY-TWO COMPANIES**

<table>
<thead>
<tr>
<th>Score</th>
<th>Companies</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Per Cent</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>15.6</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>18.8</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3.1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>6.2</td>
</tr>
<tr>
<td>1</td>
<td>18</td>
<td>56.3</td>
</tr>
</tbody>
</table>

**Technical Scores, Total.** By applying the equation found at page 67, Chapter VI, to the raw technical scores awarded to each company on the basis of the effectiveness of methods used to evaluate inside and outside programs and teaching methods, we arrive at an average combined technical score of 11.9 for the group. By multiplying this result by 2.2, the number needed to convert the maximum raw technical score to sixty (the maximum procedural score), and dividing that by forty-one, the maximum number of companies in any single technical category, we arrive at a converted average technical score for the group of 26.2 points out of a possible maximum score of sixty.
This means that, as a whole, companies in the sample made technical scores equal to 43.7% of the maximum possible technical score.

The distribution of individual companies in the sample by converted technical scores is contained in Table XLI.

**TABLE XLI**

DISTRIBUTION OF CONVERTED TECHNICAL SCORES OF FORTY-ONE COMPANIES

<table>
<thead>
<tr>
<th>Score Range</th>
<th>Companies</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Per Cent</td>
<td></td>
</tr>
<tr>
<td>56-60</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>51-55</td>
<td>1</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>46-50</td>
<td>1</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>41-45</td>
<td>2</td>
<td>4.9</td>
<td></td>
</tr>
<tr>
<td>36-40</td>
<td>3</td>
<td>7.3</td>
<td></td>
</tr>
<tr>
<td>31-35</td>
<td>6</td>
<td>14.6</td>
<td></td>
</tr>
<tr>
<td>26-30</td>
<td>4</td>
<td>9.8</td>
<td></td>
</tr>
<tr>
<td>21-25</td>
<td>7</td>
<td>17.2</td>
<td></td>
</tr>
<tr>
<td>16-20</td>
<td>14</td>
<td>34.1</td>
<td></td>
</tr>
<tr>
<td>11-15</td>
<td>2</td>
<td>4.9</td>
<td></td>
</tr>
<tr>
<td>6-10</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1-5</td>
<td>1</td>
<td>2.4</td>
<td></td>
</tr>
</tbody>
</table>

We see again (to a lesser extent than was true for scores of outside and inside program evaluation techniques considered separately) a cluster of
scores above the mean and a somewhat larger cluster below the mean. In the next chapter the significance of the clustering of scores will be examined and, if possible, explained.

Concluding Statement. The preceding mass of statistics (which admittedly makes for dull reading) can be distilled to form a composite picture of the technical effectiveness of the average company's evaluation plan.

The average company which evaluates formal management development programs appears to be concerned primarily with whether the programs are acceptable to the participants and the degree to which the programs affect job performance. It is markedly less concerned about the amount of information transmitted by the programs to the participants. The lack of emphasis on the "transmission of information" criterion is, in the author's opinion, a serious defect inasmuch as it ignores one of the most important ways by which the weaknesses of any program can be identified and remedied.

The evaluation techniques which the average company uses to measure achievement of the "acceptability" criterion are quite adequate. Likewise, when it decides to measure achievement of the "transmission of information" criterion, it usually selects adequate instruments for this purpose; however, more likely than not, it will make no effort to measure the effectiveness of programs in terms of this criterion. It usually makes an effort to see how programs have affected job performance; however, the tools it selects for this purpose are frequently inadequate.

The average company often takes more pains when it is evaluating inside programs than it does when evaluating outside programs. It will use a wider variety of evaluation techniques and those techniques are likely to be more effective than the ones it will use to evaluate outside programs.
The average company usually makes an effort to evaluate teaching methods. However, the techniques selected usually will do no more than tell whether a particular method meets a predetermined standard. They will not, as a rule, show which of two or more methods is superior for teaching a given subject matter.

To conclude: Those who have charge of management development programs in the average company do an excellent job of determining whether programs are acceptable and display much activity aimed at showing the results achieved by the programs (however, the activity often involves the use of such imprecise techniques that in most cases they will not really show to what extent a given program has improved work performance); but they neglect those techniques which will disclose areas where programs or teaching methods can be improved.

At the risk of reading too much into the data, the author would say that the typical training department is overly concerned with seeing that its programs are highly acceptable and that others are led to believe that the programs improve work performance and that this overemphasis often leads to the neglect of those kinds of evaluations which produce improvements in programs and teaching methods. The foregoing is not true, of course, of every company. A good many make earnest attempts, not only to see whether programs are a success, but to see where they may be improved. However, this does not appear to be the case for the average company.
CHAPTER X

FINAL SCORES AND RELATIONSHIPS BETWEEN VARIOUS FACTORS AND ACCEPTANCE AND FINANCIAL STABILITY OF PROGRAMS

In this chapter we shall attempt to draw together in common relationship many of the isolated facts which have come to light in the preceding sections of the thesis. First, we shall examine the combined effectiveness scores for evaluational plans of companies in the sample. Next we shall see what relation, if any, exists between the employment of persons with advanced degrees on responsible positions in management development, the use of effective evaluation plans and the acceptance and financial stability of formal management development programs. Third, we shall determine what sort of relationship exists between the effectiveness of evaluational plans and the acceptance and stability of formal management development programs. Finally, we shall see what connections exist between recruitment policies followed by management development departments, the effectiveness of evaluational plans and formal program acceptance and stability.

Relationship Between Procedural and Converted Technical Scores. Companies which scored in the top range for procedural scores tended strongly to earn high converted technical scores. Those who scored in the lower ranges for procedural scores tended strongly to earn low converted technical scores. Those who scored in the intermediate ranges for procedural scores tended to score in the intermediate converted technical score ranges, but not necessarily in accordance with their standing in the procedural score ranges. The following,
Table XLII, summarizes this information.

<table>
<thead>
<tr>
<th>Procedural Score Range</th>
<th>Number of Companies</th>
<th>Average Technical Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>56 - 60</td>
<td>14</td>
<td>34.5</td>
</tr>
<tr>
<td>51 - 55</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>46 - 50</td>
<td>4</td>
<td>21.8</td>
</tr>
<tr>
<td>41 - 45</td>
<td>3</td>
<td>28.0</td>
</tr>
<tr>
<td>36 - 40</td>
<td>6</td>
<td>26.3</td>
</tr>
<tr>
<td>31 - 35</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>26 - 30</td>
<td>6</td>
<td>19.7</td>
</tr>
<tr>
<td>21 - 25</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16 - 20</td>
<td>6</td>
<td>20.9</td>
</tr>
<tr>
<td>11 - 15</td>
<td>1</td>
<td>28.0</td>
</tr>
<tr>
<td>1 - 10</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The average combined score (procedural score plus converted technical score) for forty companies was 67.8 points out of a possible 120 points or 56.5% of the maximum possible score.

The distribution of combined scores indicates a clustering of scores above and below the mean with the cluster above the mean being slightly smaller than the one below the mean. The clustering effect, however, is not as pronounced as it was for the distributions of procedural and converted-technical scores.
considered separately. Table XLIII shows the distribution of combined scores.

**TABLE XLIII**

**DISTRIBUTION OF COMBINED SCORES FOR FORTY COMPANIES**

<table>
<thead>
<tr>
<th>Combined Score Ranges</th>
<th>Companies</th>
</tr>
</thead>
</table>
|                      | Number    | Per Cent  
| 106 - 110           | 1         | 2.5        
| 101 - 105           | 1         | 2.5        
| 96 - 100            | 1         | 2.5        
| 91 - 95             | 3         | 7.5        
| 86 - 90             | 2         | 5.0        
| 81 - 85             | 6         | 15.0       
| 76 - 80             | 2         | 5.0        
| 71 - 75             | 1         | 2.5        
| 66 - 70             | 3         | 7.5        
| 61 - 65             | 3         | 7.5        
| 56 - 60             | 1         | 2.5        
| 51 - 55             | 3         | 7.5        
| 46 - 50             | 7         | 17.5       
| 41 - 45             | 1         | 2.5        
| 36 - 40             | 2         | 5.0        
| 31 - 35             | 1         | 2.5        
| 26 - 30             | 0         | 0          
| 21 - 25             | 2         | 5.0        
| 1 - 20              | 0         | 0          |
Relationship Between Advanced Degrees and Other Factors. The combined scores for companies which reported the percentage of persons with advanced degrees holding responsible positions in management development extended from the score ranges 21-25 to 106-110. There was a definite progression in the average percentage of persons holding advanced degrees as the scores went from the lowest score range to the highest score range. The companies which fell in the combined score range 21-50 filled, on the average, 18.3% of their more important positions in management development with persons holding advanced degrees. The companies which fell in the combined score range 51-85, filled 19.1% of these positions with persons holding advanced degrees. Companies which fell in the score range 86-110, used persons with advanced degrees to fill 29.1% of their more important management development positions.

The relationship between percentage of degrees and scores, however, is an erratic one. Many of the companies with a high percentage of persons holding advanced degrees did not score significantly above the mean scores for all companies and, in fact, scored below the mean scores in a number of instances. Other companies with a relatively low percentage of persons holding advanced degrees scored very high. There is little evidence of systematic progression in scores based upon the percentage of persons holding advanced degrees. Table XLIV shows this in detail.
Table XLIV

DISTRIBUTION OF SIXTEEN COMPANIES BY PERCENTAGE OF ADVANCED DEGREES AND PROCEDURAL, CONVERTED-TECHNICAL, AND COMBINED SCORES

<table>
<thead>
<tr>
<th>Percentage With Advanced Degrees</th>
<th>Number</th>
<th>Procedural (Mean = 46.6)</th>
<th>Converted Technical (Mean = 27.5)</th>
<th>Combined (Mean = 74.1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>91 - 100</td>
<td>2</td>
<td>40</td>
<td>18.7</td>
<td>68.7</td>
</tr>
<tr>
<td>81 - 90</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>71 - 81</td>
<td>2</td>
<td>60</td>
<td>32.5</td>
<td>92.5</td>
</tr>
<tr>
<td>61 - 70</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>51 - 60</td>
<td>2</td>
<td>60</td>
<td>28.6</td>
<td>88.6</td>
</tr>
<tr>
<td>46 - 50</td>
<td>1</td>
<td>30</td>
<td>19.8</td>
<td>49.8</td>
</tr>
<tr>
<td>41 - 45</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>36 - 40</td>
<td>1</td>
<td>50</td>
<td>39.6</td>
<td>89.5</td>
</tr>
<tr>
<td>31 - 35</td>
<td>2</td>
<td>45</td>
<td>26.4</td>
<td>71.4</td>
</tr>
<tr>
<td>26 - 30</td>
<td>1</td>
<td>15</td>
<td>30.8</td>
<td>45.8</td>
</tr>
<tr>
<td>21 - 25</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16 - 20</td>
<td>1</td>
<td>60</td>
<td>44.0</td>
<td>104.0</td>
</tr>
<tr>
<td>11 - 15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6 - 10</td>
<td>2</td>
<td>60</td>
<td>24.2</td>
<td>84.2</td>
</tr>
<tr>
<td>1 - 5</td>
<td>2</td>
<td>30</td>
<td>22.0</td>
<td>52.0</td>
</tr>
</tbody>
</table>
It appears that the most important factor involved in earning high scores is not the percentage of persons holding advanced degrees used to fill the more important management development positions, but instead, is whether anyone with an advanced degree is used to fill any of the more important positions. When the average scores for all companies who have persons with advanced degrees are compared with those for companies who do not have persons with advanced degrees, it is found that the former group scores significantly higher in all respects than the latter. Table XLV shows this relationship in detail.

**TABLE XLV**

**COMPARISON OF AVERAGE SCORES OF SIXTEEN COMPANIES HAVING PERSONS WITH ADVANCED DEGREES WITH THOSE OF NINETEEN COMPANIES HAVING NO PERSONS WITH ADVANCED DEGREES**

<table>
<thead>
<tr>
<th>Companies</th>
<th>Number</th>
<th>Average Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Procedural</td>
</tr>
<tr>
<td>All</td>
<td>35</td>
<td>42.1</td>
</tr>
<tr>
<td>With Degrees</td>
<td>16</td>
<td>46.6</td>
</tr>
<tr>
<td>Without Degrees</td>
<td>19</td>
<td>38.8</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td>7.8</td>
</tr>
</tbody>
</table>

In terms of percentages, companies with persons holding advanced degrees scored 20% higher in procedural scores, 14.6% higher in converted-technical scores and 17.8% higher in combined scores than companies without persons holding advanced degrees.
This difference in scores does not appear to have any effect upon the acceptance of formal management development programs by top management or the financial stability of such programs. The average acceptance score (with one equaling least acceptable and and six, most acceptable) was 4.7 and the average financial stability score (with one equaling least stable and six, most stable) was 3.4 for both companies with and without persons holding advanced degrees.

Relationship Between Effective Evaluational Plans and Program Acceptance and Stability. Those companies which scored in the upper one-fifth and one-half ranges for procedural scores showed a moderate tendency towards higher acceptance of formal management development programs by top management. The programs of those which scored in the upper one-fifth showed a slight tendency to be more financially stable than those which scored in the remaining four-fifths. These data are summarized in Table XLVI.

TABLE XLVI

ACCEPTANCE SCORES OF THIRTY-NINE COMPANIES AND FINANCIAL STABILITY SCORES OF THIRTY-SEVEN COMPANIES BY PROCEDURAL SCORE STANDINGS

<table>
<thead>
<tr>
<th>Group</th>
<th>Acceptance Score</th>
<th>% Above Lower Group</th>
<th>Stability Score</th>
<th>% Above Lower Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper One-Fifth</td>
<td>5.0</td>
<td>8.7</td>
<td>3.4</td>
<td>3.0</td>
</tr>
<tr>
<td>Lowest One-Fifth</td>
<td>4.6</td>
<td>-</td>
<td>3.3</td>
<td>-</td>
</tr>
<tr>
<td>Upper One-Half</td>
<td>4.8</td>
<td>4.4</td>
<td>3.6</td>
<td>0</td>
</tr>
<tr>
<td>Lower One-Half</td>
<td>4.6</td>
<td>-</td>
<td>3.6</td>
<td>-</td>
</tr>
<tr>
<td>Average Scores</td>
<td>4.7</td>
<td>3.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The position of companies in the converted-technical score standings had little to do with the acceptance of management development programs. In fact, the two are so unrelated that those companies which scored in the highest and lowest fifths in converted-technical score standings had the same average acceptance scores and the average acceptance scores of both were above the mean for all acceptance scores. However, it appeared to have some effect upon the stability of programs. Those companies which scored in the upper one-fifth showed somewhat greater stability than those in the lowest one-fifth. Nevertheless, this is only a very slight tendency as attested to by the fact that both the upper and lower fifths were below the mean stability score and that the program stability of companies scoring in the upper one-half in converted technical score standing was actually three per cent lower than that of companies scoring in the lower one-half. Table XLVII summarizes these data.

TABLE XLVII

ACCEPTANCE SCORES OF THIRTY-NINE COMPANIES AND FINANCIAL STABILITY SCORES OF THIRTY-SEVEN COMPANIES BY CONVERTED-TECHNICAL SCORE STANDINGS

<table>
<thead>
<tr>
<th>Group</th>
<th>Acceptance</th>
<th>Stability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Score</td>
<td>% Above Lower Group</td>
</tr>
<tr>
<td>Upper One-Fifth</td>
<td>4.9</td>
<td>0</td>
</tr>
<tr>
<td>Lowest One-Fifth</td>
<td>4.9</td>
<td>-</td>
</tr>
<tr>
<td>Upper One-Half</td>
<td>4.7</td>
<td>0</td>
</tr>
<tr>
<td>Lower One-Half</td>
<td>4.7</td>
<td>-</td>
</tr>
<tr>
<td>Average Scores</td>
<td>4.7</td>
<td>-</td>
</tr>
</tbody>
</table>
The reader will recall that combined scores represent the effectiveness of the complete evaluational plan considered in all its relevant aspects. There was found to exist a strong relationship between high combined scores and program acceptance and stability. The relationship between combined scores and program acceptance is quite evident at all combined score levels. However, it is evident between combined scores and program stability only at the highest and lowest combined score levels. Here we find that companies which scored in the highest one-fifth for the combined score standings scored significantly higher, both above the program stability mean for all companies and the program stability mean for companies scoring in the lowest one-fifth of the combined scores standings. Conversely, those companies which scored in the lowest one-fifth of the combined score standings also scored significantly below the program stability mean for all companies. It should be noted, however, that there were no significant differences in stability scores between companies which ranked in the upper half of the combined score standings and those which ranked in the lower half. These data are summarized in Table XLVIII.
TABLE XLVIII

ACCEPTANCE SCORES OF THIRTY-NINE COMPANIES AND
FINANCIAL STABILITY SCORES OF THIRTY-SEVEN
COMPANIES BY COMBINED SCORE STANDINGS

<table>
<thead>
<tr>
<th>Group</th>
<th>Acceptance</th>
<th>Stability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Score</td>
<td>% Above Lower Group</td>
</tr>
<tr>
<td>Upper One-Fifth</td>
<td>5.3</td>
<td>12.8</td>
</tr>
<tr>
<td>Lower One-Fifth</td>
<td>4.7</td>
<td>-</td>
</tr>
<tr>
<td>Upper One-Half</td>
<td>4.8</td>
<td>4.4</td>
</tr>
<tr>
<td>Lower One-Half</td>
<td>4.6</td>
<td>-</td>
</tr>
<tr>
<td>Average Scores</td>
<td>4.7</td>
<td></td>
</tr>
</tbody>
</table>

A further indication that the financial stability of programs depends upon
the effectiveness of evaluation plans appears when the average financial sta-
bility score and average combined effectiveness score for inside programs are
compared with those for outside programs.

The average procedural score for outside programs was 15.45 (Tables
XXVI and XXVII, pages 97 and 98) and the average converted technical score was
9.46 (page 116). This gives an average combined score for outside programs of
24.91 out of a possible forty points. The average procedural score for inside
programs was 16.06 and the average converted technical score was 10.56. This
gives an average combined score of 26.62 out of a possible forty points for
inside programs.

Companies in the sample were asked whether, in times of financial
retrenchment, expenditures for outside programs are cut the same, more, or less than inside programs. If we assign a value of one to companies which reported cutting inside and outside programs to the same degree, a value of two to those which reported cutting inside programs (or outside programs, depending upon which is being considered) less, and a value of zero to those which cut inside programs (or outside programs, depending again upon which is being considered) more, we find that inside programs have an average stability score of .98 out of a possible score of two and outside programs have an average stability score of .75 out of a possible score of two. Thus, we have further evidence to support the proposition that the better the evaluation plan, the more stable the program.

**Recruitment Policies.** Companies which expressed as first choice persons with work experience in profession-related (management development-related) areas, tend to have a higher percentage of persons with advanced degrees holding responsible positions in management development and higher acceptance scores than do companies which expressed as first choice persons with work experience in business-related areas. The latter group, however, has on the average higher stability and combined scores, than the former. These data are summarized in Table II.
TABLE II

THIRTEEN COMPANIES LISTING AS FIRST CHOICE WORK EXPERIENCE IN PROFESSION-RELATED AREAS COMPARED WITH FIFTEEN COMPANIES LISTING AS FIRST CHOICE WORK EXPERIENCE IN BUSINESS-RELATED AREAS BY ACCEPTANCE, STABILITY, AND COMBINED SCORES AND PERCENTAGE OF ADVANCED DEGREES.

<table>
<thead>
<tr>
<th>Type of Work Experience</th>
<th>Number</th>
<th>Per Cent Degrees</th>
<th>Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Acceptance</td>
</tr>
<tr>
<td>Profession Related</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teaching</td>
<td>1</td>
<td>60.0</td>
<td>5.0</td>
</tr>
<tr>
<td>Industrial Relations</td>
<td>7</td>
<td>46.7</td>
<td>4.7</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>2</td>
<td>40.0</td>
<td>5.0</td>
</tr>
<tr>
<td>Training</td>
<td>3</td>
<td>12.5</td>
<td>4.8</td>
</tr>
<tr>
<td>Business Related</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management</td>
<td>7</td>
<td>17.5</td>
<td>3.9</td>
</tr>
<tr>
<td>Technical</td>
<td>3</td>
<td>31.5</td>
<td>5.3</td>
</tr>
<tr>
<td>Communications</td>
<td>5</td>
<td>40.0</td>
<td>5.0</td>
</tr>
</tbody>
</table>

Average, Profession Related (P) 30.2  4.8  3.3  63.1
Average, Business Related (B) 24.0  4.5  3.5  82.6
Difference ("P" - "B") 6.2  0.3  -0.2  -19.5
Companies which mentioned persons holding degrees in profession-related areas as first choice for recruitment tend to have a lower percentage of persons with advanced degrees holding responsible positions in management development and higher stability scores than do companies which mentioned persons holding degrees in business-related areas as their first choice for recruitment. However, the latter group earned a higher average combined score and the same average acceptance score as the former. These data are summarized by Table L.
TABLE I

THIRTEEN COMPANIES LISTING AS FIRST CHOICE DEGREES IN PROFESSION-RELATED AREAS COMPARED WITH EIGHT COMPANIES LISTING AS FIRST CHOICE DEGREES IN BUSINESS-RELATED AREAS BY ACCEPTANCE, STABILITY, AND COMBINED SCORES AND PERCENTAGE OF ADVANCED DEGREES

<table>
<thead>
<tr>
<th>Type of Degree Preferred</th>
<th>Number of Degrees</th>
<th>Per Cent Degrees</th>
<th>Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Acceptance</td>
</tr>
<tr>
<td>Profession Related</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>2</td>
<td>16.5</td>
<td>5.0</td>
</tr>
<tr>
<td>Industrial Relations</td>
<td>4</td>
<td>62.0</td>
<td>4.8</td>
</tr>
<tr>
<td>Psychology</td>
<td>5</td>
<td>25.0</td>
<td>4.6</td>
</tr>
<tr>
<td>Liberal Arts</td>
<td>2</td>
<td>0</td>
<td>4.0</td>
</tr>
<tr>
<td>Business Related</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business</td>
<td>4</td>
<td>62.5</td>
<td>4.2</td>
</tr>
<tr>
<td>Technical</td>
<td>4</td>
<td>20.8</td>
<td>5.0</td>
</tr>
</tbody>
</table>

Average, Profession-Related (P) 67.5 4.6 3.9 67.5
Average, Business Related (B) 41.7 4.6 2.9 69.9
Difference ("P" - "B") 15.8 0 1.0 -2.4
The relationships between acceptance and stability and recruitment policies are peculiar ones. We find, for example, that companies which prefer to hire persons with profession-related work experience have a higher proportion of persons with advanced degrees in profession-related areas than do companies which prefer to hire persons with business-related work experience. From this we might infer that the former group has higher combined scores and higher acceptance and stability scores than the latter. This does not prove to be the case. The former group has lower combined and stability scores than the latter, but higher acceptance scores.

The same type of relationship exists between preferences in academic backgrounds and acceptance and stability scores. Those companies which list as their first choice for recruitment persons holding degrees in profession-related fields have a lower proportion of persons with advanced degrees in profession-related fields working for them than do companies which prefer to recruit persons with degrees in business-related fields. From this one might infer that the former has lower combined, lower acceptance and lower stability scores than the latter. The former, although it actually has a lower average combined score, has the same average acceptance score and a higher average stability score than the latter.

It is possible to speculate about the causes for this, although, with the data at hand, it is not possible to prove anything. The key to the matter may lie in the fact that management development departments which are profession oriented, i.e., preferred persons with profession-related degrees and work experience, uniformly earned lower combined scores, but the same or higher acceptance scores than did those which are business-oriented. This may indicate two things: First, that the tendency to evaluate programs is an outgrowth of the
business-oriented attitude rather than the profession-oriented attitude. Sec-
ondly, that the profession-related attitude, being permitted by the company yet
being foreign to the generally business-oriented attitude of top management,
creates acceptance of its programs because it, although not understood, is it-
self accepted. To put it another way, the profession-oriented attitude acts to
create a high degree of acceptance for formal management development programs
on the rather homy basis of "I don't know what he's doing, but (because I like
him) I'm sure he knows how to do it." In short, the "halo" effect is at work.
However, when matters become serious--that is to say, when budgets must be cut
--the profession-oriented department (because it does less evaluating) cannot
advance solid proof that its programs are worth more to the company than they
cost whereas the business-oriented department usually can. Accordingly, the
programs of the former suffer more than the latter when an economy drive is in
process. Once the economy drive has ended, the profession-oriented department
regains its acceptance (which perhaps it never lost, but merely could not prove
it deserved) and things revert to the normal state.

Concluding Statement. There is no systematic, progressive relationship
between the percentage of persons who hold advanced profession-related degrees
employed by a given company, and the effectiveness of that company's evalua-
tional plan. However, there is a significant difference in combined scores
between companies which employ persons holding advanced degrees in profession-
related fields and those which do not. This fact seems to account for the
clustering of procedural, technical and combined scores at points above and
below the mean. Thus, the first null hypothesis advanced in Chapter II, page
14, is rejected. However, the second null hypothesis advanced in Chapter II,
page 15, was proved when no relationship was found to exist between acceptance
and stability scores and the fact that a company does or does not fill important management development positions with persons holding advanced degrees.

The acceptance and stability scores of companies seem to be related directly to the overall effectiveness of their evaluational plans. Those companies which earned the highest combined scores also had the highest acceptance and stability scores. Those companies with the lowest combined scores also had the lowest acceptance and stability scores. However, at the intermediate ranges for combined scores, little difference was found between the upper and lower scoring groups.

The tendency for the stability of programs to depend upon the effectiveness of evaluation plans was given further confirmation by the fact that combined scores and financial stability for inside programs were higher than those for outside programs. Considering all the evidence together, it appears conclusive that effective evaluational plans make for greater acceptance of programs among top management and for greater financial stability of programs. Accordingly, the third null hypothesis advanced in Chapter II, page 15, stands disproved.

It is rather difficult to make anything out of the data on profession-oriented versus business-oriented recruitment policies. It appears that the tendency to evaluate may inhere in the business-oriented attitude more than it does in the profession-oriented attitude. On the other hand, programs presented by business-oriented departments are not so readily accepted as those presented by profession-oriented departments. A possible reason for this is that since persons with business-oriented attitudes are better known to top management, it feels less uneasy about criticizing their work. However, the tendency to evaluate stands the business-oriented management development department in
good stead during economic crises. At such times it can answer questions about
the actual economic value of programs to the company with more assurance than
can its profession-oriented counterpart.
The process of management development has become increasingly rationalized in the past few years. Formal training of one kind or another has been used with growing frequency to supplement experience in sharpening managerial skills. Today large sums of money are spent by corporations to support a wide variety of formal programs used both to prepare managerial personnel for greater responsibilities and to increase their effectiveness on their present jobs.

There is some evidence to indicate that formal management development programs initially gained uncritical acceptance. However, as their newness wore off, they became subject more frequently to attack from a number of quarters. Today a number seriously question whether anything is received in return for the sums of money spent by corporations on formal management development programs.

This hardening attitude has engendered, in turn, a desire to show that programs actually produce beneficial results. The upshot is that the desirability of evaluating programs for the purpose of determining whether they have achieved their objectives has generally been accepted by critics and defenders, alike.

The benefits to be derived from carefully designed evaluations are many--more effective training, more economical use of training resources, and
increased confidence in training activities, to mention a few. However, one writer, using an ultimate criterion of development (the classical "educated man") as the standard for measuring program success, cast doubt upon the possibility of securing meaningful data from evaluations. While his point might be well taken if that particular criterion is to be used, evaluations, nevertheless, will yield meaningful data if they are made in terms of the limited, concrete objectives which usually form the purpose for any given formal program.

The subject of evaluation of formal programs is a live one—one which is relevant to much of today's controversy about the value of management development. The author undertook this study with the hope of shedding some light upon the extent and effectiveness of program evaluation plans as they are presently used in a substantial portion of industry.

Problem. The problem chosen as the subject for study was to determine the extent to which the one hundred largest U. S. manufacturing corporations (from the standpoint of sales) evaluate the effectiveness of formal management development programs conducted or participated in by them and the effectiveness of techniques used by them to evaluate programs.

Among the secondary problems investigated were the interrelationships between the degree of professionalization (employment of persons holding advanced degrees in work-related fields) in management development departments, the effectiveness of evaluational plans, the degree of acceptance of programs by top management, and the financial stability of programs.

Methodology and Limitations. A six-page questionnaire was constructed and sent to each company in the sample. The questionnaire was accompanied by a cover letter which gave the reasons for the study and a sheet of instructions
which specified how and by whom the questionnaire should be completed. Returns of questionnaires from companies in the sample formed the primary source of data for the study.

Although the sample covered by the study includes companies which are among the largest and most prosperous found in this country, it is neither particularly large nor representative. Therefore, conclusions drawn by the study cannot be said to hold true for all of American corporations. However, it can be said with reasonable assurance that the effectiveness of evaluational plans found in the sample is as high, if not higher, than the effectiveness of similar plans found in the rest of industry.

The questionnaire method of collecting data poses its own peculiar difficulties—there is no way in which to insure response from each company and it is difficult to formulate questions in such a way as to make them free from all ambiguity. Moreover, in order to keep the size of the questionnaire within manageable limits it was necessary in many instances to use questions which produced approximate rather than exact information.

Characteristics of Effective Evaluational Plans. In order to determine the relative effectiveness of evaluational plans, it is first necessary to decide upon what features are essential to effectiveness. Effectiveness has two aspects—procedural and technical. An evaluational plan, no matter how technically perfect, is not effective unless it covers most or all of the formal management development programs with which the company is involved. Conversely, no matter how complete the coverage of the evaluational plan, it is not effective if it employs faulty evaluational techniques.

When the concept of procedural effectiveness is subjected to further analysis it breaks down into two parts, viz., coverage and use. An
evaluational plan is not procedurally complete unless it covers all types of programs with which the company is involved and those teaching methods which are within the control of the company. It is procedurally not complete, even though the maximum coverage has been achieved, if the data from evaluations are not used as the primary basis for making administrative decisions about continuing, modifying or discontinuing programs and teaching methods.

The characteristics of technically effective evaluational plans were considerably more difficult to isolate. Finally, it was reasoned that if the purpose of formal management development programs is to improve work performance, then the success of programs should be measured in terms of the degree to which they improve work performance. Accordingly, the tools used to evaluate programs must be capable of determining the effect a given program has on work performance. Next it was reasoned that formal programs accomplish their effects by transmitting information to the participants which, if applied, will result in improved work performance. It follows that evaluation tools should be capable of determining the amount of information transmitted by programs to the participants. Finally, it was reasoned that if programs are not acceptable to the participants, they will produce little good no matter how relevant their material. It follows, then, that evaluation tools should also be able to determine the acceptability of programs to the participants.

These three technical criteria do not overlap. Although the over-all success of a program may be determined by using the "improvement in work performance" criterion, this will not tell the observer what portions of the program can stand improvement. This can be worked out only in terms of the criteria of "transmission of information" and "acceptability." If there is
generally poor acceptability, the causes for this usually will appear in evaluations made in terms of the "acceptability" criterion. If certain information is not being presented adequately, this will appear in evaluations made in terms of the "transmission of information" criterion. However, evaluations made in terms of these two criteria will not give the observer any certain indication that programs have had a favorable impact upon work performance. Programs may, as shown by a number of writers, be highly acceptable and may communicate information satisfactorily, but still fail to improve work performance. Thus while the three criteria are interdependent, they stand as separate and distinct entities.

In general terms, an evaluational plan at this point is considered to be procedurally complete if; (1) it covers all programs in which the company is involved, (2) covers teaching methods where they are within the control of the company, and (3) uses evaluation data as the basis for important administrative decisions about programs and teaching methods. An evaluation plan is considered technically complete if evaluation tools used are capable of; (1) measuring the effect produced by the program upon work performance, (2) measuring the amount of information transmitted by the program to participants, and (3) measuring the acceptability of the program to the participants.

After taking the foregoing into account, it was decided that evaluation plans should be rated according to their procedural and technical completeness and that both factors should be given equal weight in establishing the effectiveness of evaluation plans.

**Techniques for Evaluating Program Effectiveness.** No matter which technique is used to measure the improvement in job performance caused by a program, it is essential to use an untrained control group as the basis for
determining improvement in the trained group. Without a control group, not even the most refined technique can distinguish between results produced by the program and those caused by the mere passage of time.

It is not necessary, however, to use control groups in order to measure attainment of the "transmission of information" criterion. The training program, since it exists to transmit certain information in concentrated form, constitutes a monopoly of that information for all practical purposes. The possibility that factors outside the control of the program may act to furnish the participants with significant amounts of information of the type which forms the subject matter of the program is too remote to justify going to the trouble of selecting and testing a control group.

Control groups have no place in techniques used to determine the acceptability of programs. Control groups, as that term is used here, are not involved in programs; hence they can have no opinion one way or another about the acceptability of a program. Consequently, they cannot be used as a basis of comparison for groups which, by virtue of attending a program, have the experience necessary to form an opinion of its acceptability.

Four techniques are used rather widely by industry to evaluate management development programs. These are:

1. Questioning program participants to find out their opinion of the program.
2. Giving attitude or achievement tests to the participants before beginning and after completing the program, with or without giving the same tests to untrained control groups.
3. Comparing the participants' actual work performances before beginning the program with their performances after completing
the program, with or without making the same comparisons with untrained control groups.

4. Comparing ratings made by the participants' superiors, peers or subordinates of the participants' work performance before beginning and after completing the program, with or without making similar comparisons between ratings made of untrained control groups.

No matter how adroitly opinions are solicited, the fact remains that this procedure will tell the observer little, if anything, about what the participants learned or how what they learned affected their work performance. Accordingly, the first technique is of little use in measuring attainment of the "improvement in work performance" and the "transmission of information" criteria. However, it goes to the very heart of what determines the acceptability of a program--namely, the opinions the participants have formed about the program. Therefore, it is eminently suitable for measuring attainment of the "acceptability" criterion.

The various testing techniques tell rather little about the acceptability of programs. However, they disclose a great deal about the amount of information transmitted to participants by the program. Their usefulness in this respect is undiminished even when control groups are not used.

If tests are to be used to measure the effect of programs upon work performance, they must be used in conjunction with control groups. However, questions were raised about their adequacy to measure improvement in work performance under any circumstances. The major objection to their use for this purpose is the fact that test scores may change without corresponding changes taking place in work performance. Experimental findings by
Fleishman, Levin and Butler all tend to substantiate the objection.

After weighing the evidence, it was finally concluded that test results may be used as indicators of changes in work performance provided; (1) the tests are used in conjunction with control groups and (2) the tests have been validated on criterion groups performing the same or similar types of work as that which the development program seeks to improve.

Changes in grievance, absentee, turnover, suggestion and transfer rates may serve as direct indicators of changes in work performance for managerial employees. The use of these, and similar indicators are universally acknowledged to be the best method for determining the actual effect of formal management development programs upon work performance. However, their utility is confined largely to measuring attainment of the improvement in work performance criterion. Although they will give a very accurate account of a program's over-all success, they are not capable of discriminating between the transmission of information and acceptability criteria and cannot be expected to disclose areas where the program can be improved.

The fourth technique—that of comparing ratings made of the participants' work performances by their superiors, peers, or subordinates before beginning and after completing the program—is, by its very nature, confined to measuring the changes effected in work performance by programs. It may not be used to determine the extent to which programs have satisfied the acceptability and transmission of information criteria. Like other methods of measuring changes in work performance, it should be used in conjunction with control groups.

The rating technique is not as satisfactory as those which directly measure changes in work performance. In essence, it is a judgment about changes in work performance and, like all judgments, is subject to the individual
standards, bias, and personal opinions of the judges. However, when rating
standards have been thoroughly explained to the raters and are understood more
or less alike by them, the rating technique is reasonably satisfactory for use
in determining whether a given program has succeeded or failed.

Evaluating Teaching Methods. There is close to universal agreement be­
tween authorities in the field of education that the worth of various teaching
methods is best determined by the amount of information conveyed to and re-
tained by students. The primary criterion for the evaluation of teaching
methods, then, is student learning.

Two general approaches may be taken towards evaluating teaching methods.
First, a given method may, without reference to other methods, be judged in
accordance with a pre-established standard of adequacy. Secondly, a method
used to teach certain subject matter may be compared with another method used
to teach the same subject matter. The second approach is usually considered
the superior of the two. It goes beyond the first by showing not only which
methods are adequate, but also which one of a number of methods is best suited
for teaching a given subject.

The author found that four types of evaluational techniques are used
generally to determine the effectiveness of teaching methods. These are;
(1) testing, (2) ratings by instructors, (3) ratings by participants, and
(4) ratings by outside observers. In practice, the second and third tech­
niques are not used with the comparative approach. The first and fourth
techniques, however, are often used with both the comparative and the pre-
established standard approaches.

The use of test scores to evaluate the effectiveness of teaching methods
has long been recognized as one of the more adequate evaluational techniques.
Its nature is such that it may be used with equal facility to determine whether a given method has attained a pre-established standard of adequacy or which of two or more methods is best suited for teaching a given subject. Of all the evaluational techniques, it is the one used most frequently for comparisons between teaching methods. Its use for this purpose is thoroughly documented in educational literature.

The use of ratings by instructors as the basis for evaluating the effectiveness of teaching methods is a less desirable technique than the use of test scores. Frequently, no effort is made to see that all instructors understand the rating procedure alike. As a result, each has his own opinion about what is needed for a given method to reach a pre-established standard of adequacy. If the relative effectiveness of two or more methods is to be determined on the basis of this technique, the difficulties are multiplied. The resulting comparative ratings may tend to amount to little more than an aggregation of personal opinions.

Evaluating teaching methods by using outside observers trained in the use of standardized rating guides, is a technique which overcomes or minimizes to a great extent the difficulties noted in ratings by instructors. Most often, there is adequate consistency (rater reliability) between ratings made by different observers of the same technique. However, since the technique consists of passing judgment upon what appears to have been learned, rather than measuring directly what has been learned, it is considered somewhat less satisfactory than techniques involving the use of tests.

The least satisfactory evaluational technique is that which involves the use of ratings by students. The student is not a trained observer. Even when he makes the effectiveness of teaching methods a paramount consideration in
the formation of his judgment, his evaluation necessarily rests upon private standards of effectiveness. The greatest objection, however, is that the student may not use the effectiveness of teaching methods as the primary basis for his judgment. This objection is not at all farfetched. A sizable number of published reports disclose instances where such judgments were based on the entertainment qualities of teaching methods rather than their effectiveness.

Rating Scale. The author devised a scale for the purpose of rating the procedural and technical effectiveness of evaluational plans used by companies in the sample. Each company was assigned a procedural score of sixty from which deductions were made for failure to evaluate specified types of programs and teaching methods, or for failure to evaluate repeated programs or use evaluational data as the basis for administrative decisions about changing, continuing or discontinuing programs and teaching methods.

In determining the relative weight to be given for failure by a company to cover specified types of programs or teaching methods in its evaluational plan, it was decided that the rank order of deductions would be, from least to most, as follows:

1. Short ad hoc programs.
2. Long ad hoc programs.
3. Short repeated programs.
4. Long repeated programs.
5. Teaching methods.

The rank order of deductions, from least to most, for failure to use evaluational data as the basis for administrative decisions is:

1. Short repeated programs.
2. Long repeated programs.
3. Teaching methods.

Equations were devised to compensate for variations which might be introduced by companies having in their management development program less than the maximum possible number of elements (programs, teaching methods, possibility of using evaluation data, etc.) which can be included in such programs.

In constructing the equations, it was assumed that the smaller the variety of management development program elements, the more serious becomes failure to include any given element in an evaluation plan.

The techniques used by companies to evaluate inside and outside programs and teaching methods were rated separately in accordance with their ability to measure attainment of the three general program criteria. The order of program evaluation techniques in relation to the improvement in work performance criterion, is from least to most valuable, as follows:

1. Questioning participants, testing participants without use of control groups, and rating without use of control groups.

2. Measurement of work performance without the use of control groups.

3. Testing or rating participants with the use of control groups.

4. Measurement of work performance with the use of control groups.

The order of program evaluation techniques, (from least to most valuable) according to their capacity to measure attainment of the transmission of information criterion, is as follows:

1. Questioning participants and rating or measuring participants' work performance, with or without the use of control groups.
2. Testing participants either with or without the use of control groups.

The order of program evaluational techniques, (from least to most valuable) according to their capacity to measure attainment of the acceptability criterion, is as follows:

1. Testing, rating or measuring work performance, with or without the use of control groups.

2. Questioning participants.

Techniques used to evaluate teaching methods were ranked in accordance with their ability to measure student learning. Their rank order, from least to most effective, is as follows:

1. Ratings by students, no comparisons.

2. Ratings by instructors, no comparisons.

3. Ratings by outside observers, no comparisons.

4. Test scores, no comparisons.

5. All types of comparative techniques.

A number of equations were devised for the purpose of combining technical scores assigned to techniques used separately in the evaluation of inside and outside programs and teaching methods. In addition, a conversion factor was used to give equal weight to technical scores when they were combined with procedural scores to produce a score representing the over-all effectiveness of evaluational plans.

General Findings. Sixty-nine per cent of the companies in the sample responded to the questionnaire. Of these, the great majority report that they have both inside and outside formal management development programs.

Companies in the sample tend to be involved with a greater variety of
inside programs than outside programs. Most rely to a greater extent upon outside concerns for ad hoc programs than they do upon their own training organizations. Conversely, most companies rely more heavily upon their own training organizations to handle repeated programs than they do upon outside concerns.

Most companies with formal management development programs have plans for evaluating the programs. For the most part, evaluation plans cover nearly all types of programs with which the companies are involved. Despite this, a substantial minority of companies with evaluation plans do not use evaluation data as the primary basis for administrative decisions about continuing, modifying or discontinuing programs.

Approximately three-fourths of the companies with inside programs said that they evaluate teaching. Somewhat more than half of this group reported using evaluation data as the primary basis for making administrative decisions about continuing, discontinuing or modifying teaching methods.

Companies tended to exclude from their evaluation plans a higher proportion of all types of outside programs, except long repeated programs, than they did types of inside programs. Moreover, even where the various types of outside programs participated in by a given company were included in the evaluation plan, the company tended to rely less upon evaluation data from outside repeated programs for use as the basis for administrative decisions than it did upon data taken from evaluations of inside repeated programs.

The average combined score for companies in the sample was 67.8 out of a maximum possible score of 120 points or 56.5% of the maximum possible score. The distribution of combined scores for individual companies clustered at points above and below the mean with the cluster above the mean being slightly
A positive relationship appears to exist between procedural and converted technical scores. Companies who scored in the top range for procedural scores (56-60 points) tended strongly to earn high converted technical scores. Those who fell in the lower ranges for procedural scores tended strongly to earn low converted technical scores. Those who scored in the intermediate ranges for procedural scores tended to earn intermediate converted technical scores, but not necessarily in accordance with their standings in the procedural score ranges.

This relationship indicates that the technical and procedural aspects of evaluational plans are interdependent. That is, if the plan is procedurally sound, it will tend to be technically sound and, conversely, if it is procedurally deficient it will tend to be technically deficient. This, in turn, is a strong indication that the technical and procedural effectiveness of any evaluational plan depends upon either a single factor or upon a number of closely related (and probably mutually dependent) factors.

There were wide variations from company to company in the proportion of persons with advanced academic degrees in work-related fields who held responsible positions in management development departments. A little more than half of the companies were without persons holding advanced degrees. Of those companies with persons holding advanced degrees, the proportion of advanced degrees ranged all the way from five per cent to one hundred per cent.

The recruitment policies followed by companies seeking new personnel to fill positions in management development departments also varied widely. Four companies require advanced degrees in work-related fields as a prerequisite to employment. The remainder do not have this requirement. Companies in the
sample were evenly divided between those who prefer new employees to have academic degrees in business-related fields and those who prefer them to have academic degrees in management development-related fields. However, those companies who prefer to hire persons with work experience in management development-related fields outnumbered slightly those who prefer to hire persons with experience in business-related fields.

Of those companies which answered questions relating to acceptance of formal management development programs by top management, a sizable majority indicated that most of their top management believe that management development programs contribute more to the welfare of the company than they cost.

The financial stability of management development programs appears to be equal to or better than that of other phases of company activity. Companies in the sample indicated that when expenses are cut generally, management development program budgets usually are not cut so heavily as other items of expense. However, within the training budget, there is a strong tendency to make larger cuts in expenditures for outside management development programs than in those for inside management development programs.

The acceptance of formal management development programs by members of top management appears to be directly related to the financial stability of formal management development programs. Those companies which reported the highest proportion of acceptance, also reported the greatest degree of stability. Those which reported the least amount of acceptance, reported the least degree of financial stability. Those which reported acceptance falling in the middle range, reported financial stability in the middle range.

**Procedural Effectiveness.** The average procedural score for companies which furnished answers to all questions bearing upon procedure was rather low--42.8
points out of a possible sixty points (71.3% of maximum possible score). The range for procedural scores extended from fifteen points to sixty points. The distribution of procedural scores was skewed on the high side. One-third of the companies earned procedural scores falling in the fifty-six to sixty score range. The remaining companies were fairly evenly distributed among all but the very lowest score range.

There is a very strong tendency for companies to include within the coverage of evaluation plans all types of management development programs in which they are involved. Out of a possible score of ten points, companies in the sample averaged 9.6 points on program coverage. The tendency for companies to evaluate teaching methods within their control is not quite so strong. Companies in the sample averaged 7.4 points out of a possible score of ten points for coverage of teaching methods.

The tendency among companies to use evaluation results as the basis for making administrative decisions about continuing, modifying or discontinuing specific programs and teaching methods is considerably weaker than the tendency to evaluate. Of those companies which reported evaluating short and long repeated outside programs, only 54.4% and 64.3%, respectively, said that they use evaluation data from these sources as the primary basis for administrative decisions. The tendency to use for this purpose data taken from evaluations made of inside programs is somewhat, but not greatly, stronger than it is for outside programs. Of those companies which reported evaluating short and long repeated inside programs, 62.5% and 65.6%, respectively, said that data from these sources are used as the primary basis for administrative decisions. A somewhat higher proportion (68.8%) of the companies which evaluate teaching methods indicated that data taken from evaluations are used as the
primary basis for administrative decisions about particular teaching methods.

**Technical Effectiveness of Evaluation Methods.** The average company earned a converted technical score of 26.2 points out of a maximum possible score of sixty points (43.7% of the maximum possible converted technical score). Distribution of converted technical scores for individual companies clustered at points above and below the mean. The cluster above the mean is somewhat smaller than the one below the mean.

Companies in the sample earned 4.3 points out of a possible nine points for techniques used to evaluate outside programs and 4.7 points out of a possible nine points for techniques used to evaluate inside programs. There are indications that more pains are taken with evaluations of inside programs than outside programs. A wider variety of evaluational techniques is used and, as the scores indicate, those techniques tend to be more effective.

The types of evaluation techniques reported, indicate that companies who evaluate formal management development programs are concerned primarily with (1) whether the program is acceptable to the participants and (2) the degree to which the program affects job performance. They appear to be significantly less concerned about the amount of information transmitted by the program to the participants. The lack of emphasis on this point is, in the author's opinion, a serious defect inasmuch as it means that one of the most important means by which the weaknesses of programs may be identified is being neglected.

Evaluation techniques used by companies in the sample are quite adequate for measuring attainment of the "acceptability" criterion. Techniques used to measure attainment of the "transmission of information" criterion likewise are adequate; however, more likely than not, no effort will be made to measure the
effectiveness of programs in terms of this criterion. Usually some effort is made to determine the extent to which programs affect job performance; however, the evaluational tools selected for this purpose frequently prove to be inadequate.

The majority of companies make efforts to evaluate teaching methods within their control. However, the evaluation techniques usually employed will do no more than tell whether a particular teaching method meets a predetermined standard. They will not, as a rule, show which of two or more methods is superior for use with a given subject.

**Evaluation Effectiveness; Acceptance and Stability.** The effectiveness of evaluation plans has a definite influence upon the degree of acceptance of programs by top management and the financial stability of programs. The influence is felt most strongly at the upper and lower ranges of effectiveness. Companies whose evaluation plans ranked in the upper one-fifth of combined score standings had stability and acceptance scores which were nearly thirteen per cent higher than those of companies whose evaluation plans ranked in the lowest one-fifth of combined score standings.

Generally speaking, evaluation effectiveness exerted slightly more influence upon acceptance scores than upon stability scores. Companies which ranked in the upper one-half of combined score standings had, on the average, acceptance scores 4.4% higher than those in the lower one-half. However, there was no difference in average stability scores between companies which fell in the upper and lower halves of combined score standings.

One may conclude that evaluational effectiveness has a direct and very nearly proportionate effect upon acceptance of management development programs by top management. However, it has little influence upon the financial
stability of programs except at the very highest and lowest levels of effec-
tiveness.

Effect of Advanced Degrees. There is no progressive relationship between
the percentage of persons holding advanced work-related degrees employed by a
given company and the effectiveness of that company's evaluation plan. How-
ever, there is a significant difference in the effectiveness of evaluation
plans of companies which employ persons (regardless of number) holding advanced
degrees in work-related fields and those who do not. The former group aver-
ages higher procedural, technical, and combined scores than the latter. It
appears that the split between companies which employ persons holding advanced
degrees and those who do not, accounts for, in all probability, the clustering
of various types of scores at points above and below the mean.

As mentioned earlier, evaluation plans must be either very good or very
poor before they can exert a substantial influence upon the acceptance and
financial stability of management development programs. The average effective-
ness of evaluation plans of both companies which do and do not employ persons
holding advanced degrees does not lay at either extreme. Consequently, it
was found that the acceptance and financial stability scores for the two
groups were the same.

Effect of Recruitment Policies. Companies which mentioned persons holding
degrees in profession-related (management development-related) areas as their
first choice for recruitment, tend to have considerably less effective evalu-
ation plans than do companies which mentioned persons holding degrees in
business-related areas as their first choice for recruitment. Despite this,
the former tends to have somewhat higher acceptance scores than the latter
although the latter exceeds the former in financial stability scores.
Companies which mentioned as their first choice for recruitment persons with work experience in profession-related areas tend to have slightly more effective evaluation plans than do companies who mentioned as their first choice for recruitment persons with work experience in business-related areas. Both have the same acceptance scores. However, management development programs of the former group show greater financial stability than those of the latter group.

It is rather difficult to make anything out of the data on profession-oriented and business-oriented recruitment policies followed by various management development departments. The major reason for this appears to be that the number of companies which gave sufficient answers to permit comparisons between recruitment policies and evaluation effectiveness, acceptance and financial stability was too small to prevent one or two atypical companies from distorting the picture. However, despite conflicts and distortions, it appears that the tendency to evaluate inheres more in the business-oriented attitude (taken here as the motive force behind business-oriented recruitment policies) than it does in the profession-oriented attitude. The tendency to evaluate stands the business-oriented management development department in good stead during times of economic crisis by enabling it to answer questions about the economic value of the programs to the company with more assurance than its profession-oriented counterpart. As a consequence, its programs usually have greater financial stability than those of profession-oriented departments. On the other hand, programs given by business-oriented departments are not as widely accepted as those given by profession-oriented departments. A possible explanation for this is that since persons with business-oriented attitudes are more familiar to top management, the latter
feels more free to criticize their work.

**Conclusion.** The employment of one or more persons with advanced academic degrees in work-related areas on responsible positions in management development appears to be a key factor in the design and use of effective evaluation plans. In addition there are some indications that using persons with advanced degrees who have had work experience in other areas of management tends to increase further the effectiveness of evaluation plans.

Whether or not a given management development department is business-oriented or profession-oriented also appears to influence the effectiveness of evaluation plans. Those departments which are business-oriented definitely tend to use more effective evaluation plans.

The evaluation plans of companies in the sample thoroughly cover inside and outside programs of various types and teaching methods. As a rule, techniques used to evaluate programs can determine, quite adequately, the degree to which the programs are acceptable to the participants. However, there appear to be two major weaknesses in evaluation plans. Although most companies make serious efforts to determine what effect programs have had upon work performance, the techniques used for this purpose are, for the most part, methodologically weak. The second weakness is that most companies neglect techniques which, by isolating weak points in programs or comparing the relative effectiveness of different teaching methods, make it possible to improve systematically programs and teaching methods. The rather prevalent distrust of using evaluation data as the basis for important decisions about the fate of programs and teaching methods appears to grow naturally out of the use of inadequate evaluational techniques.

It should, nevertheless, be kept in mind that the use of formal programs
for management development is a relatively new thing. Until such time as it becomes accepted as a normal facet of corporate activity, those who have charge of programs will naturally concentrate more upon justifying the existence of programs and making them acceptable to participants than upon systematically improving the quality of programs. However, with the passage of time, we may hope to see a diminishing of this defensive attitude and the rapid growth of one which seeks to find new and better ways of improving the quality of programs.
BIBLIOGRAPHY

I. PRIMARY SOURCES

A. BOOKS


B. ARTICLES


----. "Evaluating the Results of Supervisory Training," Personnel, XXX (January 1957), 362-370.


"Companies Sound Off on Executive Training," Dun's Review and Modern Industry, LXVII (June 1956), 89-90.

Colyer, Luther. "A Comparison of Two Methods of Teaching Biology at the College Level," Science Education, XLIV (February 1960), 52-58.


Maier, Norman R. F. "An Experimental Test of the Effect of Training on Discussion Leadership," Human Relations, VI (Fall 1953), 161-173.


Savitt Morris A. "Is Management Training Worth While?," Personnel, XXXIV (September-October 1957), 79-82.


II. SECONDARY SOURCES

A. BOOKS


Hayes, Benjamin R., M. E. Broom and Mathilde Hardaway. Tests and Measurements in Business Education. Cincinnati, 1940.


B. ARTICLES


Thisdell, Robert A. "Why Not Measure Training Results?," Journal of the American Society of Training Directors, XIII (October 1959), 9-12.


APPENDIX I

CORPORATIONS INCLUDED IN SAMPLE

-A-

Allied Chemical Corporation
Allis - Chalmers Manufacturing Company
Aluminum Company of America
American Can Company
American Cyanamid Company
American Motors Company
American Radiator and Standard Sanitary Corporation
American Smelting and Refining Company
American Tobacco Company
Anaconda Company
Armco Steel Company
Armour and Company
Atlantic Refining Company

-B-

Bendix Corporation
Bethlehem Steel Corporation
Boeing Airplane Company
Borden Company
Borg-Warner Corporation
Burlington Industries Incorporated

-C-

Campbell Soup Company
Caterpillar Tractor Company
Chrysler Corporation
Cities Service Company
Coca - Cola Company
Colgate - Palmolive Company
Continental Can Company Incorporated
Continental Oil Company
Corn Products Company
Crown Zellerbach Corporation

-D-

Deere and Company
Douglas Aircraft Company Incorporated
Dow Chemical
E. I. DuPont de Nemours Company

-E-

Eastman Kodak Company

-F-

Firestone Tire and Rubber Company
Ford Motor Company

-G-

General Dynamics Corporation
General Electric Corporation
General Foods Corporation
General Mills Incorporated
General Motors Corporation
General Telephone and Electronics Corporation
General Tire and Rubber Company
B. F. Goodrich Company (The)
Goodyear Tire and Rubber Company
W. R. Grace and Company
Gulf Oil Corporation

-I-

Inland Steel Company
International Business Machines Corporation
International Harvester Company
International Paper Company
International Telephone and Telegraph Company

-J-

Jones and Laughlin Steel Corporation
-K-
Kennecott Copper Corporation

-L-
Lockheed Aircraft Corporation

-M-
Martin Company
Minnesota Mining and Manufacturing Company
Monsanto Chemical Corporation
John Morrell and Company

-N-
National Cash Register Company
National Dairy Products Corporation
National Lead Company
National Steel Corporation
North American Aviation Incorporated

-O-
Olin Mathieson Chemical Company
Owens-Illinois Glass Company

-P-
Phillips Petroleum Company
Pittsburgh Plate Glass Company
Procter and Gamble Company
Pure Oil Company

-R-
Radio Corporation of America
Ralston Purina Company
Raytheon Company
Republic Steel Corporation
R. J. Reynolds Tobacco Company
-S-
St. Regis Paper Company
Shell Oil Company
Sinclair Oil Corporation
Singer Manufacturing Company
Socony Mobil Oil Company
Sperry Rand Corporation
Standard Oil Company of California
Standard Oil Company of Indiana
Standard Oil Company of New Jersey
J. P. Stevens Company Incorporated
Sun Oil Company
Sunray Mid-Continent Oil Company
Swift and Company

-T-
Texaco Incorporated
Tidewater Oil Company

-U-
Union Carbide Corporation
United Aircraft Corporation
United Merchants and Manufacturers Company
United States Rubber Company
United States Steel Corporation

-W-
Western Electric Corporation Incorporated
Westinghouse Electric Corporation
Weyerhaeuser Company
Wilson and Company Incorporated

-Y-
Youngstown Sheet and Tube Company (The)
APPENDIX II

I am writing you about a study I am making on techniques used by industry to evaluate the results of formal management development programs and the effectiveness of training methods used in the programs.

The purpose for the study is twofold. First, it will be used as a thesis written to fulfill part of the requirements for the degree of Master of Science in Industrial Relations at Loyola University. Secondly, the study should be of some benefit to my employer, the Illinois Central Railroad, which is presently considering initiating a formal management development program.

The study represents one of the first comprehensive efforts made to ascertain and evaluate techniques actually used to measure the effectiveness of formal management development programs and training methods. The conclusions drawn by the study should be of value to persons interested in the control aspects of management development programs.

The study is under the direction of my advisor at Loyola University and is being carried out with the help and active encouragement of the Illinois Central Railroad. The best method of securing data for a study of this type would be the interview. However, because of the scope of the study, that method presents insurmountable problems. Accordingly, the attached questionnaire represents the most practicable method available. Information given in reply to the questionnaire will be held in confidence and the names of companies participating in the study will not appear in the final paper.

Would you please have the questionnaire completed and returned to me at the address shown above. Thanking you in advance for your time and trouble,

Very truly yours,
INSTRUCTIONS

1. It is desired that person answering questionnaire either be in charge of management development or be the immediate subordinate of person in charge of management development.

2. The symbol "X" is to be used in answering multiple-choice questions. More than one item appearing in such questions may be checked.

3. Please staple attachments directly to completed questionnaire.

4. Definitions:

A. Control group--group composed of persons not participating in management development program. It should be similar in most other respects to participant group. It is used as a basis for determining improvement in participant group.

B. Management development program--formal program, akin to class work, designed to improve managerial performance at present level or to prepare individuals for greater managerial responsibilities.

C. Management development, responsible positions--other than clerical positions. Includes relatively complex technical duties or supervisory responsibilities over technical positions. Positions must be directly concerned with planning, administering or conducting management development programs.

D. Participants--individuals who attend management development programs in role of students.

E. Program, ad hoc--special purpose program which is given only once.

F. Program, inside--management development program conducted for managerial employees by respondent company.

G. Program, outside--management development program attended by managerial employees of respondent company, but conducted by organization outside the company, i.e., AMA seminars, university programs, etc.

H. Program, repeated--program attended by successive groups of participants. Differs from ad hoc program in that it is conducted more than once.

I. Test, reliability--the consistency of scores obtained by the same individuals on different occasions, either before attending a program or after attending a program but not before and after attending a program, on a given test.

J. Test, validity--the degree to which the test actually measures what it is supposed to measure. Determining the validity of a test usually involves comparing test results with an independent, external criterion of that which the test is designed to measure or predict.

5. Please attach whatever descriptive literature your company may have prepared dealing with management development programs conducted by it.
I. GENERAL INFORMATION

1. Name of Company___________________________________________________________.

2. Does your company conduct or participate in management development programs? Yes____ No_____ (If the answer to this question is "no", the remaining questions should not be answered).

3. When your company recruits new employees to work with management development programs, emphasis is placed on securing men with work experience in the following fields (please list in order of preference):
   A. ________________________________________________________________
   B. ________________________________________________________________
   C. ________________________________________________________________
   D. ________________________________________________________________
   E. ________________________________________________________________

and with academic degrees in the following fields (please list in order of preference):
   A. ________________________________________________________________
   B. ________________________________________________________________
   C. ________________________________________________________________
   D. ________________________________________________________________
   E. ________________________________________________________________

4. Is the possession of an advanced degree a prerequisite to employment on responsible positions with management development programs? Yes____ No_____. If answer to this question is "yes", please indicate in which of the following areas an advanced degree is required: Personnel____; Industrial Relations____; Education____; Sociology____; Psychology____; Others (please specify)____.

5. Approximately what percentage of employees presently holding responsible positions in management development programs have earned advanced degrees in the work related areas enumerated in question 4?
   ____%
II. OUTSIDE MANAGEMENT DEVELOPMENT PROGRAMS OTHER THAN THOSE LEADING TO FORMAL DEGREES

1. Does your company participate in outside management development programs?  
   Yes  No

2. If answer to question 1 is "yes", please answer the remaining questions in this section. Check which of the following types of outside programs are participated in by your company:

   A. ______ Ad hoc programs requiring twenty hours or less of participants' time.
   B. ______ Repeated programs requiring twenty hours or less of participants' time.
   C. ______ Ad hoc programs requiring more than twenty hours of participants' time.
   D. ______ Repeated programs requiring more than twenty hours of participants' time.
   E. ______ Others (please specify) ____________________

3. Have any of the foregoing types of programs been evaluated in order to determine whether they have achieved their objectives?  Yes  No

4. If answer to question 3 is "yes", please check appropriate blanks below.

<table>
<thead>
<tr>
<th>Type of Program</th>
<th>Frequently Evaluated</th>
<th>Occasionally Evaluated</th>
<th>Never Evaluated</th>
<th>Evaluation results primary determinant of whether program will be continued.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes  No</td>
<td>Yes  No</td>
<td>Yes  No</td>
<td>Yes  No</td>
</tr>
<tr>
<td>Short ad hoc</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short repeated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long ad hoc</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long repeated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Which of the following methods are used to evaluate outside programs:

   A. ______ Examination of participants on subject matter of program before commencing and after completing program.

   B. ______ Same technique as described in "A" except that improvement in performance of participant group is compared with improvement of non-participating control group over same time span.
C. Measurement of participants' job performance after completion of program and comparing results with job performance before commencing program.

D. Same technique as described in "C" except that improvement in performance of participant group is compared with improvement in performance of non-participating control group over same time span.

E. Questioning participants to determine their opinion of value of program.

F. Questioning participants' superiors about whether they feel participants have benefited from program.

G. Questioning participants' peers about whether they feel participants have benefited from program.

H. Others (please specify) __________________________

NOTE: This is one of the most important questions in the questionnaire. Please attach, if possible, copies of actual evaluational reports and memoranda describing evaluational methods and techniques. If tests are used, please describe the methods by which they were checked for reliability and validity.
III INSIDE MANAGEMENT DEVELOPMENT PROGRAM

1. Does your company conduct inside management development programs?  
   Yes _____ No _____

2. If answer to question 1 is "yes", please answer the remaining questions in this section. Check which of the following types of inside programs are conducted by your company:

   A. _______ Ad hoc programs requiring twenty hours or less of participants' time.
   B. _______ Repeated programs requiring twenty hours or less of participants' time.
   C. _______ Ad hoc programs requiring more than twenty hours of participants' time.
   D. _______ Repeated programs requiring more than twenty hours of participants' time.
   E. _______ Others (please specify) ____________________________________________

3. Have any of the foregoing types of programs been evaluated in order to determine whether they have achieved their objectives?  Yes _____ No _____

4. If answer to question 3 is "yes", please check appropriate blanks below.

<table>
<thead>
<tr>
<th>Type of Program</th>
<th>Frequently Evaluated</th>
<th>Occasionally Evaluated</th>
<th>Never Evaluated</th>
<th>Evaluation results primary determinent of whether program will be continued</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short ad hoc</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short repeated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long ad hoc</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long repeated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Which of the following methods are used to evaluate inside programs:

   A. _______ Examination of participants on subject matter of program before commencing and after completing program.

   B. _______ Same technique as described in "A" except that improvement in performance of participant group is compared with improvement in performance of non-participating control group over same time span.
C.____ Measurement of participants' job performance after completion of program and comparing results with job performance before commencing program.

D.____ Same technique as described in "C" except that improvement in performance of participant group is compared with improvement in performance of non-participating control group over same time span.

E.____ Questioning participants to determine their opinion of value of program.

F.____ Questioning participants' superiors about whether they feel participants have benefited from program.

G.____ Questioning participants' peers about whether they feel participants have benefited from program.

H.____ Others (please specify) ____________________________________________________________

NOTE: This is one of the most important questions in the questionnaire. Please attach, if possible, copies of actual evaluational reports and memoranda describing evaluational methods and techniques. If tests are used, please describe the methods by which they were checked for reliability and validity.

6. Are educational techniques, i.e., case studies, lectures, incidents, role playing, etc., evaluated for effectiveness? Yes _______ No _______.

7. If answer to question 6 is "yes", which of the following methods are used to determine the effectiveness of educational techniques:

A.____ Examination of participants of subject matter of program.

B.____ Examination of participants and comparison of results with those derived from examinations on same subjects given participants subjected to different techniques.

C.____ Questioning participants to determine their reactions to various techniques.

D.____ Others (please specify) ____________________________________________________________

NOTES: Please attach, if possible, copies of actual evaluational reports and memoranda describing their use.

8. Are evaluations used as primary criteria in determining whether a particular method will be continued, modified or discontinued? Yes____ No______.
IV STABILITY OF PROGRAMS

1. The following proportion of top-level company officers believe that management development programs contribute more to the company's welfare than they cost: All _____ Most _____ More than half _____ Half _____ Less than half _____ Few _____ None _____.

2. When it is necessary for your company to reduce expenditures, approximately to what degree are expenditures for management development cut relative to other company activities such as research, advertising, public relations, maintenance, etc.? No cut _____; much less _____; much more _____; slightly less _____; slightly more _____; the same _____.

3. When expenditures for management development are reduced, outside programs are reduced in comparison with inside programs to the following degree: The same _____; less _____; more ____. 


The thesis submitted by John F. Roberts has been read and approved by three members of the faculty of the Institute of Social and Industrial Relations.

The final copies have been examined by the director of the thesis and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the thesis is now given final approval with reference to content, form, and mechanical accuracy.

The thesis is therefore accepted in partial fulfillment of the requirements for the Degree of Master of Social and Industrial Relations.

Date: 8-2-63
Signature of Advisor: [Signature]