

Forced Choice Evaluation Of A Training Program

KARL R. KUNZE

Literature of the past few years gives evidence of a growing desire among industrial training people to measure the effects of their efforts. Before-and-after comparisons of trainee knowledge and behavior; changes in production rate, work quality, cost realization, morale, etc.; trainee evaluations; and other methods are now in use and contributing to our insight into effectiveness of training programs.

There is also a dissatisfaction among training men with the present measuring devices, and this is wholesome. Obviously there is room for improvement in all of them. On the other hand, it is wholesome that devices are being employed, however inaccurate, because rough measurement is usually better than none at all. The steel rule was used before the micrometer; in fact it must have contributed to the development of the micrometer.

This article proposes another kind of measurement of training programs, one also far from perfect, but one that partially overcomes weaknesses of some devices now in use.

Forced-choice scales are being used in industry today in performance rating of employees. Proponents of this type of scale contend, and with good justification, that the "halo" effect, leniency-severity tendencies and other influences indigenous to the usual scales, are minimized. The forced-choice method has now been employed in personality tests,¹ to make them less transparent and less subject to both deliberate and involuntary slanting. The writer knows of one use of the forced-choice method in evaluating training programs² and expects that other efforts are under way. However, this method does not yet seem to be given the attention it deserves by training men.

KARL R. KUNZE is manager of the Training Department of the Lockheed Aircraft Corp., California Div. He attended Boston University and the Harvard Graduate School of Arts and Sciences, and holds an A.M. degree in psychology. Member of Industrial Relations Research Association, American Psychological Association, National Vocational Guidance Association, and ASTD. Published articles in the industrial relations field and instructs at the California Institute of Technology, Industrial Relations Section, and the University of Southern California, Extension Div.

-
1. The Gordon Personal Profile is an example.
 2. Blansfield, M. G., "University Executive Development and the Trainer," *Journal of the ASTD*, Vol. 12, No. 2, February, 1958.

When used as an employee performance rating, the forced-choice scale usually consists of blocks of statements. Some statements are favorable in overtone, others unfavorable, and some may be of a neutral nature. The rater is asked to select, in each block, the one statement that is most descriptive or characteristic and the one statement that is least descriptive of the person being rated. The rater finds himself forced to choose one favorable descriptive item and leave others unchecked. He must also indicate which of two or more negative items is least characteristic of the ratee, with the awareness that the remaining items will be considered as more characteristic. The rater knows that values have been established for the items but does not know what they are.

The forced-choice method will not be described in any greater detail here. Good articles are available on the subject.³ It is suggested here that the forced-choice method is singularly appropriate to the measurement of training programs. A training program has many facets and can be described in many different ways, like an employee's work performance, or someone's personality.

This method reduces some of the inaccuracies inherent in present measurement methods.

The writer has evaluated several of his training programs with the forced-choice method. Following is the general procedure that has been used.

When they had met sufficiently to obtain an impression of the training program's nature and content, participants were asked to think of statements that could be considered descriptive of the course being given. In some classes, buzz groups were used to increase the variety of responses.

Duplications were then eliminated, ambiguities corrected, and items added when necessary. The trainees were then asked to indicate the degree of importance of each item if, later on, it should prove to be descriptive of the course. In the compilation of the items, three degrees of importance, little, moderate, and very, were given in columnar form. Trainees were asked to check the column appropriate to each item.

Using the trainees' responses, weighted scores were then assigned the items as shown in Figure 1.

FIGURE 1

	Most Characteristic		Least Characteristic	
	Favorable	Unfavorable	Favorable	Unfavorable
Of high importance	+5	-5	-5	+5
Of medium importance	+3	-3	-3	+3
Of low importance	+1	-1	-1	+1

The final rating scale was then prepared. Two favorable and two unfavorable items were placed in each block.

The sequence was randomized in terms of the degree of importance, and favorableness or unfavorableness of the items.

3. See *Forced-Choice Performance Reports—A Modern Merit Rating Method*, M. W. Richardson, *Rating Employee and Supervisory Performance*, American Management Assoc., N. Y., 1950.

One of the scales used appears in Figure 2. This scale was administered in two similar programs so that results could be compared. In the instructions, raters were asked (1) to select one statement in each block that was most characteristic of the training program under consideration and to indicate their choice with an "X" in the answer column; and (2) to "X" the one statement least char-

acteristic of the program. An example was given of a block of four statements with two of them checked.

Before discussing results obtained in the applications of this method described above, it should be noted that this type scale yields two kinds of findings: (1) a distribution of scores and (2) responses to individual items.

FIGURE 2
A FORCED-CHOICE RATING OF A TRAINING PROGRAM

		(Answer Column)		
		MOST		LEAST
1.	A Instructor had subject matter well planned.	()	A	()
	B The sequence of topics was meaningful.	()	B	()
	C Program was slow to build up steam.	()	C	()
	D There was too much emphasis on production and not enough on office and technical training problems.	()	D	()
2.	A The course tended toward one segment of industry.	()	A	()
	B The course provided information of high value to me.	()	B	()
	C Group reaction to the course was good.	()	C	()
	D The course content was too theoretical in nature.	()	D	()
3.	A The material and ideas presented were practical and usable.	()	A	()
	B Time was poorly apportioned to each subject.	()	B	()
	C The course did not meet its stated objectives.	()	C	()
	D The course provided a good exchange of ideas.	()	D	()
4.	A The supporting information was insufficient.	()	A	()
	B The method (presentation-discussion) used was effective.	()	B	()
	C The order of topics could have been improved upon.	()	C	()
	D The course provided an opportunity to compare practices of several companies.	()	D	()
5.	A The course was a complete waste of time.	()	A	()
	B The scope of the program was too extensive.	()	B	()
	C A high interest was maintained.	()	C	()
	D The course provided new insights and approaches to supervisory training problems.	()	D	()

		(Answer Column)	
		MOST	LEAST
6.	A The outline was followed closely.	()	A ()
	B The course assumed more background in training than the participants possessed.	()	B ()
	C The time spent on each subject was sufficient.	()	C ()
	D The course failed to meet the training needs of present-day industry.	()	D ()
7.	A Instructor's presentation was disorganized.	()	A ()
	B The time spent on each subject was sufficient.	()	B ()
	C The course met its stated objectives.	()	C ()
	D Some topics could have been omitted without reducing the value of the course.	()	D ()
8.	A The group was homogeneous.	()	A ()
	B The supporting information was insufficient.	()	B ()
	C The presentation of material stimulated comment and controversy.	()	C ()
	D The outline was not followed closely.	()	D ()
9.	A There was a good balance between theory and practice.	()	A ()
	B Visual aids were used effectively.	()	B ()
	C The information presented was of limited value to my company.	()	C ()
	D The course did not meet its stated objectives.	()	D ()
10.	A The course content was not "pulled together" or summarized adequately.	()	A ()
	B The instructor was not very humorous.	()	B ()
	C The material presented was up to date.	()	C ()
	D The sequence of topics was meaningful.	()	D ()

As in the case of any distribution of scores, comparisons can be made of one with another. The scores obtained from one training program can be compared with another in which changes in the form or techniques of training have been instituted. Of course in such a study, the same scale would be used for both programs, although the weights might be different (inasmuch as you might be interested in the reaction to the items in terms of the values placed on them by the trainees). The distribution itself can be analyzed. The range

of scores can be compared with the total possible range of the scale. The lowest score obtainable on a scale can be ascertained by summing the scores of the favorable items of least importance (having lowest weight) in the "most" column, plus the unfavorable items of most importance in each block of four. The highest score would be just the reverse: the score of the most favorable items of most importance, plus the unfavorable items of least importance in each block.

The zero score on the scale can be used in a consideration of scores because

this represents a point of balance between favorable and unfavorable items of equal importance.

In Figure 3 are distributions of the responses of two classes held at the Industrial Relations Section of the California Institute of Technology on "Developing and Administering Training Programs.." Of the maximum range of the scale, 168 points (from -84 through zero to +84), 102 points were used. It should be noted that many low scores were obtained, especially in the course given February, 1957.

FIGURE 3
DISTRIBUTIONS OF SCORES
FROM TWO TRAINING PROGRAMS
USING A FORCED-CHOICE SCALE

	Feb. 1957 Class	Mar. 1958 Class
75		XX
70		XX
65		
60		X
55		
50		
45		X
40	XX	X
35	X	
30	XXXX	XX
25	X	
20	X	
15	X	XX
10	X	
5		XX
0	X	
-5	X	
-10	X	
-15		
-20		
-25	X	
-30	X	

The improvement in scores of the March 1958 class over those of February 1957 can be considered a change in attitude of trainees toward aspects of the training program. Of the 1957 data, 50% of the scores were below 20, whereas 31% were below this mark in 1958. This is a statistically significant difference.⁴

In the writer's opinion, an analysis of the responses to the actual items provides the most valuable information. This is done by recording the important or highly-weighted items, both favorable and unfavorable, that are checked with frequency. If a favorable item is checked in the "least" characteristic column, this obviously should be construed as a criticism. Also, if an unimportant unfavorable item is checked as "least" characteristic, then it can be assumed that a more important negative item in the block of four is more characteristic.

An analysis of the Cal-Tech returns indicated that:

Group reaction to the course was good.
The course provided information of high value.

The material and ideas presented were practical and usable.

The course provided new weights and approaches to supervisory training programs.

The material stimulated comment and controversy.

The method used (presentation-discussion) was effective.

The material was up to date.

However, the analysis also indicated that:

4. The critical ratio in this instance is 10.6 (1% level of confidence).

The supporting information was insufficient.

The scope of the program was too extensive.

The instructor was not very humorous.

The outline was not followed closely.

Some topics could have been omitted without reducing the value of the course.

The course assumed more background in training than the participants possessed.

As a concluding comment, the forced-choice scale is proposed as a means for measuring the effectiveness of training

programs. In all likelihood this form of measurement will not displace any now in use, but might prove to be a worth-while supplement to them. The forced-choice scale should prove superior to the usual questionnaire as a training measuring device.

More extensive use of the scale for this purpose, and more statistical study would be helpful. For example, the reliability of the degrees of importance of items could be looked into. The writer plans to do some of this and hopes others will be motivated to try this type of scale. The writer will give any assistance requested and would appreciate information concerning any results obtained.

An ASTD Research Committee Report on . . .

Communication And Human Relations

CLYDE S. HARTLOVE, Chairman

In the Human Relations area, I think we might use the work being done by Dr. Allen Solem of the University of Maryland.

In 1957 and again in April 1958, the Maryland Chapter has had Dr. Solem conduct Human Relations seminars and workshops. At these workshops, he has stressed FRUSTRATION; MOTIVATION; and INTERVIEWING techniques aimed at creating changes in attitudes of people.

At each of these workshops Dr. Solem has used ROLE PLAYING by dividing the 300 persons attending each program into small groups. After lecturing on the various topics, he would then issue roles which are involved in case problems out of the book *Supervisory and Executive Development—A Manual for Role Play-*

ing—by Norman Maier and Allen Solem. After practice sessions, all groups reported and the results were compiled on boards and discussions then held on changes which might have been effected during the role playing.

While the basic concept of the project is not new, I feel that the techniques used, along with the book mentioned, would be an excellent addition to any organization's program in Human Relations. It gives those participating definite exposure to the fact that for any type of Human Relations Training, there must be a change in attitude of those involved if we wish to be successful.

The Maryland Chapter is preparing a paper which will be placed in the ASTD Training Materials Exchange.