

An Evaluation Of Interdepartmental Training With Objective Tests

JOSEPH GILBERT, HENRY G. CAMPBELL and ALBERT E. OLIVER

This article reports on the development of a series of brief tests based on the content of standard training guides, and easily and effectively used to assist in evaluating quantitatively and objectively the results of interdepartmental training. Not only have the test scores significantly discriminated by a statistical test between groups taking it before training as compared to groups taking it after training, thereby unequivocally measuring the effect of training, but the tests also reveal (by an analysis of those passing and those failing specific items) areas of the course that the subjects misunderstood or otherwise adequately comprehend, so that information the

tests show was not clear to subjects after training can be emphasized in a post-training review session. Moreover, the testing procedure stimulated considerable interest and proved to be an unexpectedly motivating influence, both for the subjects and for the instructors.

This report will concern two tests constructed out of five covering the primary interdepartmental training courses out of the series for the fiscal year. Assignments to individual training groups were arranged on the basis of convenience, available time from duty and so on, before it was definitely decided to embark on the testing program, and such assignments to groups

DR. JOSEPH GILBERT is *Chief Psychologist, Richland County Mental Health Clinic, Columbia, South Carolina, and Lecturer, University of South Carolina. Prior background includes psychologist experience at Veterans Administration and other hospitals. His Ph.D. is from the University of Tennessee. Dr. Gilbert has done extensive work in test design.*

H. G. CAMPBELL, *Clinical Psychologist, is Assistant Director, Veterans Administration Hospital, Fayetteville, North Carolina. Previously with the Veterans Administration he has served as Registrar, VA Hospital Rutland Heights, Massachusetts; Providence, Rhode Island; and Boston, Massachusetts. He was also Special Assistant to Manager, VA Hospital, Newington, Conn. and Northport, Long Island, N. Y. Mr. Campbell attended Boston University and Middlesex College of Medicine & Surgery and was graduated from Beacon Institute of Podiatry. He has a private practice and during World War II was with the U. S. Army as Major, M. S. C.*

were not influenced by the clinical psychologist or by the testing program. The membership of each training group was exactly as it would have been if no testing program had been initiated. This is necessary if statistical comparisons of groups are to be meaningful.

The first test constructed by the clinical psychologist was a 29 item True-False inventory for the course in Disciplinary Procedures. To avoid test-retest effects separate groups were compared before and after training rather than individuals with themselves, since the very process of taking the same test twice will familiarize an individual with it and may substantially improve his chances of doing better on it the second time than he did the first, even if he knows no more about the subject the second time.

Student Population

Four groups took the Disciplinary Procedures course, one group each week for a period of four weeks. All groups contained both men and women and a variety of vocational positions, nurses, supervisory and non-supervisory personnel, cooks, dieticians, guards, and aides. The first and third group were taught by a woman instructor, the second and the fourth by a man.

The 29 item True-False inventory was made up from information on procedure contained in the Veterans Administration Disciplinary Procedures

training manual. It was administered to the first group taking the course on the first day before the course began. It required about 15 minutes, with everyone allowed to finish the test. Scoring was simply the number right. The second group was tested after training was completed in exactly the same way. The third group was then tested after training and the fourth group before training, following the same procedure as in the first and second group.

The results of the Disciplinary Procedures testing appears in Table 1. (PRE indicates group was tested before training, POST indicates group was tested after training):

In Table 1, N refers to the number of persons taking the test in each group, M refers to the Mean or Average score of the group and the Range is the lowest and the highest score in the group. Deviation² and SD refer to the sum of the deviations of the individual scores from the mean squared and to the standard deviation respectively. Even to the most cursory inspection, however, it is apparent that the test strikingly differentiates between groups which have taken training and those which have not, making it at once clear that training in Disciplinary Procedures was quite successful, that people who took the course knew a great deal more about the subject than people who had not taken it. The pre and post training groups almost do not overlap in range, the groups tested after training and the

ALBERT E. OLIVER is Personnel Officer, Veterans Administration Hospital, Fayetteville, N. C. He attended Illinois Wesleyan University, Bloomington, Illinois, prior to entry in Armed Services where he served approximately four and one-half years. He began his career with Veterans Administration at the VA Hospital, Marion, Illinois, in 1946. Mr. Oliver has served in various Personnel Assistant positions in VA Hospitals, Marion, Illinois, and VAH, Birmingham, Alabama.

TABLE 1
Differences Between Groups on a Test of Disciplinary
Procedures Administered Before and After Training

	GP 1 PRE	GP 2 POST	GP 3 POST	GP 4 PRE
N	13	10	11	10
M	13	20	24	14
SD	4.1	2.4	2.6	2.4
Range	6-19	17-25	19-27	10-18
Deviation ²	216	57	72	59

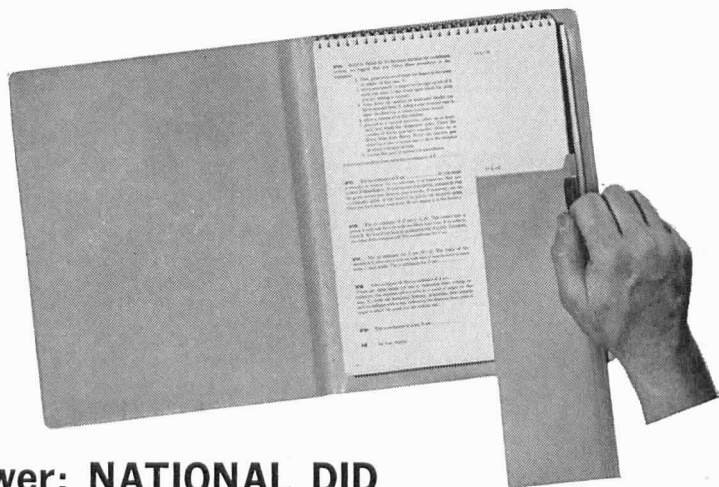
groups tested before training forming nearly separate distributions. Thus, the two groups tested *before* training, Group 1 and Group 4, had very nearly the same average score, 13 and 14 right respectively, even though Group 1 was quite variable both probably in range or ability and in previous experience with disciplinary procedures, while Group 4 was much less variable and had an unusually large number of individuals with previous experience in disciplinary procedures as it happened.

On the other hand, the two groups tested *after* training, Group 2 and Group 3, had an *average* of 20 and 24 answers correct respectively, from seven to eleven correct answers more than the two groups tested before training. In addition the *lowest* scores in Groups 2 and 3, the groups tested after training, are almost as high or higher than the highest scores in Groups 1 and 4, the groups tested before training. And the results were consistent; it was not in this instance a single occurrence of one group being compared with one other group. Rather there was a replication, an opportunity to compare two additional groups with each other and the same kind of differences were found in the same direction.

Each group tested before training was compared with each group tested after training, using Fisher's "t" test for the significance of the difference for independent small samples. The "t" for Group 1 as compared to Group 2 was 4.66; for Group 1 as compared to Group 3 it was 7.33; for Group 3 as compared to Group 4 it was 8.9; for Group 2 as compared to Group 4 it was 5.4. All of these differences between groups tested before training and tested after training were significant at better than the one percent level of confidence. The difference between the two groups tested *before* training was not significant in spite of the contrast between them in certain respects (t equals 0.67).

There was an additional significant difference between groups of particular interest because of what it indicates. This was the difference between the two groups tested *after* training, Group 2 and Group 3, smaller than any of the others (3.33) but still significant at the one percent level of confidence. Group 3, which was significantly different from Group 2 although both groups were tested after training, was taught by the woman instructor after she saw a copy of the test and specifically emphasized in her class the content

WHO DEVELOPED A NEW CONCEPT IN PROGRAMMED LEARNING AIDS FOR AS LOW AS \$1.50 PER UNIT?



Answer: NATIONAL DID

Now, from National Blank Book, comes a teaching device so simple, it can be explained to the newest learner in less than a minute. The secret of Learn-Ease® simplicity is the slide mask*—reveals or covers each program frame smoothly and easily, yet will not pull off either end of the track because of a built-in stop.

*Patents applied for

LEARN-ease UNITS ARE UNIQUE

- They are economical: the most expensive Learn-Ease is less than five dollars—and prices are as low as \$1.50 per unit in quantity
- They can be custom-designed to fit any training program—yet the most complicated mask has only one moving part
- They weigh less than two pounds (including a content load of 80 program sheets), fit easily into a briefcase or bookcase
- They are as compact in size as any standard ordinary ring binder

No bulky mechanical devices or expensive, complicated, gadgety machines. Learn-Ease units are simple, compact, rugged, can be used over and over again!

Find out how Learn-Ease can simplify your training program. Write to Programmed Learning Department T, National Blank Book Company, Holyoke, Massachusetts.



WHERE OUTSTANDING VALUE IS A MATTER OF RECORD

which the test questions covered. Group 2, on the other hand, was taught by a man who had not seen a copy of the test and did not attempt to teach specifically the content which the test questions covered.

The difference between the two groups is of considerable value in demonstrating that stressing and reviewing specific designated points is the best way of putting them across so that people do not forget them, and that if the content of a test is a good sample of the content of a course one wishes to put across, it is a worthwhile technique to teach around the correct answers to that test. However, it must be recognized that every instructor should have an equal opportunity to do this. The information which was obtained by giving the instructor of Group 3 an advantage was important, but it cannot be denied that it was unfair to the instructor of Group 2. Had the instructor of Group 2 taught the content of the test rather than the instructor of Group 3, it is not unlikely that the results would have been reversed. Actually, of course, it is apparent that both instructors did a superior job of teaching their groups the content of the course.

Item Analysis

After the results were in, the psychologist performed an item analysis to determine the percentage passing and failing each item on the test for the combined Groups 1 and 4 (before training) as compared to the combined Groups 2 and 3 (after training) to determine the hardest and easiest items on the test, as well as the items which best discriminated between the pre and post

training groups. Groups 1 and 4 (tested before training) had a total of 47 percent of the items on the test correct, while groups 2 and 3 (tested after training) had a total of 77 percent of the items on the test correct. For all groups combined there were seven items which only 25 percent of the subjects missed and there was one item which more than 75 percent of the subjects missed. There was also one reversal; on one test item the two groups tested before training (Groups 1 and 4) obtained a larger percentage of correct answers (61 percent) than did the two groups tested after training (Groups 2 and 3, 48 percent of which had this item correct).

All of the subjects taking the course next were assembled and the items of the test missed by 50 percent or more of the subjects were reviewed and the correct answer explained. This was an additional opportunity to put across both the content of the course and the philosophy behind disciplinary procedures in the V.A., taking advantage of the interest of the subjects in the outcome of the testing in which they had participated. The principal points which we sought to emphasize were the authority and responsibility of the non-voting chairman, the fact that the hearing is not legalistic in form, and that it is not punitive but fact-finding. For example, it was stressed that the first item on the test, which 68 percent of the group missed, "The *primary* purpose of the disciplinary hearing is to determine whether the defendant employee is at fault as charged," had to do with a subtle point but one which involved attitude and could determine the entire tone and course of the hearing. The

item, of course, was false; as the training manual indicates, the primary purpose of the procedure is to give the defendant employee *a chance to be heard*, with the secondary purpose being to determine the facts of the case. Some of the most frequently missed items dealt with the recommended actual procedure of the hearing as well as its fundamental objectives.

Group Meeting Test

The next course in which groups were to be compared before and after training was that of Better Group Meetings. The content by the nature of the subject matter is much more abstract than that of Disciplinary Procedures and constructing a suitable inventory was a difficult task for the psychologist, especially since he was resolved to avoid items based on inconsequential and frivolous distinctions which he felt would be irritating and unjust to subjects taking the test. He decided on a multiple choice test in which the task was to circle the number of the alternatives which best represented methods of dealing with persons who are a problem in group meetings (as argumentative, overtalkative, and so on), as indicated in the

training manual. Because of the more intangible character of the material the psychologist had less faith in this test than in the first and we felt more than a little dubious of our chances of obtaining significant differences between pre and post training groups. The test was administered to the first group before training and to the second group after training, exactly as before except that one instructor was used and there was one pre and post training group each instead of two. There were a maximum of 32 correct answers out of a total of 60 alternatives. Scoring was again the number right. The results appear in Table 2. As before, PRE indicates group was tested before training, POST indicates group was tested after training.

Thus, Group 2, the group tested after training obtains an average of six more correct answers than does Group 1, the group tested before training. At first glance the range in each group appears similar because in each the lowest score is 9. However, in Figure 1 the real difference between the distributions of the two groups is evident.

In Group 1 the majority of the scores are in the middle range, while in Group 2 the scores tend to pile up on the high

TABLE 2
Differences Between Groups on a Test of Better Group Meetings
Administered Before and After Training

	GROUP 1 PRE	GROUP 2 POST
N	16	16
M	15	21
SD	3.5	5.8
Range	9-24	9-28
Deviation ²	191	542

The
American Society of Training Directors
announces

**A New ASTD Institute
in
Programmed Instruction**

How to Use Programmed Instruction for Better Training Results

**June 25-27, 1963
Ambassador Hotels, Chicago, Illinois**

Program coverage includes these topics:

Determining Objectives and Content

Selection of Programs

Teaching Machines vs

Programmed Books

Evaluating Programmed Instruction

Practice Frame Writing

Training Programmers

and other essential information.

Registration Fee *

\$130 for ASTD members

\$155 for non-members

* Includes all materials and lunches. Ambassador Hotels reservations available at special rate of \$12.00 per day, single or \$16 per day, double.

Gordon M. Bliss, Executive Director
American Society of Training Directors
2020 University Ave., Madison 5, Wisconsin



Attendance will be limited . . . Make your reservations now

and certainly just the process of being appraised puts everyone on their toes if the procedure is fair and is not abused.

Other Applications

Can tests such as ours be used in the same way in other industrial firms and government agencies? We believe that they can if someone is willing to take the responsibility of carrying through on the testing and will keep individual scores really secret. The tests are easy to score and a mean for each group tested can be simply calculated (the sum of the scores for the group divided by the number of persons in the group taking the test). A mean difference score of five or more favoring the group taking the test after training is certainly indicative that the group tested after training has learned from the experience, providing naturally there has been no effort to manipulate the results. It is necessary for someone familiar with statistics to compute the significance of the difference, however, using a formula which takes into consideration the size of the groups, and their variance, and we believe this task is best assigned to the personnel psychologist.

Valuable as the tests are in motivating personnel and in demonstrating objectively the effectiveness of training, they

are equally valuable in revealing areas in which training for some reason is not successful. An item analysis of the kind we used, computing the percentage passing each item, is especially useful in clarifying for management and personnel ways in which training can be improved, by placing emphasis on material which appears not to be understood, or by whatever other procedure appears to be warranted. We feel that a review of the test material with the persons who took the course can be of real value with certain kinds of courses, the importance of which in the view of management justifies this arrangement. Even if this review does not appear warranted, the subjects should still be informed in general terms of the outcome of the testing program in which they participated; a brief mimeographed summary is probably adequate in most cases to satisfy the human desire for knowledge of results.

With the exception of the one experiment mentioned above of dividing a test in half and administering it before and after training, differences between pre and post training groups on *all five* tests were significant at or better than the one percent level of confidence. The other three tests are You Meet the Public by Letter, Work simplification, and Delegation of Authority.

REFERENCES

1. Adkins, Dorothy C., Primoff, E. S., McAdoo, H. L. Bridges, C. F., and Forer, B. "Construction and Analysis of Achievement Tests." Washington, U. S. Civil Service Commission, 1947.
2. Cronbach, L. J. "Essentials of Psychological Testing." New York, Harper and Brothers, 1949.
3. Diamond, Solomon. "Information and Error." New York, Basic Books, 1959.
4. Smith, G. M. "A Simplified Guide to Statistics." (Rev. ed.) New York, Rinehart and Company, 1946.