

Better Tests

Here's a procedure that's simple, yet rational

Robert C. Yelland

Many of us are inclined to bemoan the tendency of training directors to evaluate their programs in a nonprofessional manner. In so doing, the classic article by Kirkpatrick on evaluation of training programs often is given reference.¹ He lists a four-step quality hierarchy:

1. Trainee reaction
2. Achievement measure
3. Behavior on the job
4. Economic results.

There is ample evidence that attempts to evaluate training in terms of on-the-job behavior are very infrequent.² Further, most of us would agree that a cause and effect relationship in terms of economic results is almost impossible in light of our current state of knowledge. So, the best we usually do is to use an achievement measure.

Often this is in the form of a *pretest* and a *post-test*. Do we, at least, prepare "good" tests? A casual survey of members, Southern Minnesota Chapter of ASTD, indicated that many have little training in test construction. In effect, a common comment indicated that some information in the way of a "cook book" approach could be helpful.

The starting place is the intent of the course. Just what change of knowledge is desired? The training director will probably set several objectives. He or she can list these in priority order and weight them Next, write multiple-choice questions that pertain to the objectives in proportion to the weighting assigned Then institute an ongoing practice of item analysis for the questions.

Sound simple? It can be, but often isn't. The procedure that follows will try to keep it simple, yet rational.

Multiple-Choice Questions

Training directors have little trouble listing objectives, so let's consider the next point. Why multiple-choice questions. Some authorities indicate that wide knowledge of the subject areas is best measured by multiple-choice tests. There is quite a bit of evidence regarding the fallibility of the common essay questions, but that is yet another story.

Multiple-choice questions have no magic in themselves. They are no better than their reliability. Reliability, as used here, refers to the test's ability to produce the same level of score each time the test is

given to someone with the same knowledge. "How to measure that?" Give the test to the same person twice. "But then he or she would remember the question." Okay, break the test into halves, then figure out if the trainees tend to get the same score on both halves.

Use the Spearman-Brown prophesy formula to get the reliability between the scores.³ That is, take the correlation times two and divide the total by the correlation, plus one. You should seek a reliability of at least .60; .90 would be considered outstanding.

"Yes, but a correlation??? Statistics? Me?" Well, how about finding the proportion of trainees who are in the top 50 per cent of the group on both halves of the test, then use the table depicted in Figure 1.⁴

FIGURE 1.

Percentage	Correlation	Reliability
45	.95	.97
40	.81	.89
35	.60	.75
30	.31	.47
25	.00	.00

"What if the reliability is below .6?" Easy answer! POOR TEST. "What to do about a poor test?" Improve it by working with the difficulty level and discrimination index. Difficulty level . . . develop the multiple-choice questions until you find 45 per cent to 55 per cent of the training group gets each question wrong. Then, work with those questions until the trainees in the top quarter on total scores do better on each individual question than the trainees in the lower

quarter, at least 20 per cent better. After you have done these two things, you have a better question.

In working with the improvement of multiple-choice questions, here are a few hints:

1. The correct answer often tends to be the longest.
2. Correct answers tend to find a place in the middle.
3. Make sure that each of the incorrect alternatives has its share of answers.

Continue working with the test

Trained and productive people—A nation's most valuable asset.

The development of human resources is our business. Transferring skills and technologies across language and cultural barriers is our product. Telemedia is an international training firm dedicated to developing the world's manpower through a total systems approach to large international projects. For solutions to overseas training problems, call on the proven capabilities of a full service training organization—Telemedia.

See our display, Booth 525, ASTD convention or visit us at our Hospitality Suite—New Orleans, Marriott.



500 N. Michigan Avenue
Chicago, Illinois 60611 U.S.A.
(312) 644-4100 Telex: 25-4003

Telemedia, Inc.



questions and try to develop quite a few extras so you use the same ones less frequently. This will be tough to accomplish because you will find that those who work with this method tend to have only 20 per cent or so of the multiple-choice questions meet the double criterion of difficulty level and discrimination index the first time they are used.⁵

One of the really nice things about this approach is the possibility that a whole group composed of superior trainees could show as exactly that, superior. Another advantage, if the training director finds test scores, over a period of time, tend to be dropping off, then maybe something else is suffering, perhaps the presentations or the interest level of the trainees. Whatever, an investigation may be

started and probably quite a bit sooner than would otherwise be the case.

One final thought. With four-part multiple-choice questions the "chance" score is 25 per cent. A score of 25 per cent or less therefore indicates "no knowledge." This seems like a reasonably acceptable way of setting the absolute floor for a minimum passing score.

USEASTD

REFERENCES

1. Kirkpatrick, D.L., "Techniques for Evaluating Training Programs," *Journal of the American Society of Training Directors*, December, 1959, January, 1960, February, 1960.
2. Duel, Henry J., Dennis Seidman, William B. Comm and W. Eugene Temple, "A Staff Report on Evaluation of Trainee and Education in Federal Agencies," *Self and Service En-*

richment to Federal Training, U.S. Civil Service Commission, 1967, p. 421.

3. Helmstadter, G. Croft, *Principles of Psychological Measurement*. Appleton - Century - Crofts, New York, 1964.
4. Educational Testing Service, "Shortcut Statistics for Teacher-made Tests," Second Edition, 1964.
5. Notes from Psychology 270: Psychological Measurements. University of Minnesota, May, 1967, Professor Rene' Dawis.

Robert C. Yelland is manager of Management Development and Administrative Services, Midland Cooperatives, Inc., Minneapolis, Minn. He received the M.A. in Industrial Relations from the University of Minnesota. He is a member of ASTD and the American Society for Personnel Administration.

President's Banquet to Offer Mardi Gras Re-Enactment

In order to give the attendees some of the true flavor of New Orleans, the ASTD President's Banquet on May 19, 1976 will take the form of an authentic Mardi Gras Supper Dance.

During the Mardi Gras Season there are numerous street parades and many elaborate private balls, each with a theme based on History, Mythology or Literature.

The organizations which put these on are known as Krewe and all costs are borne by the members.

Presented at our banquet will be a reenactment of the Ball Masque of the Krewe of Venus. There will be a special presentation known as a "tableaux," which will depict a particular theme with the "Krewe" and "Court" in elaborate and glittering Mardi Gras costumes.

Krewe members will throw favors and specially minted coins which are called "doubloons." A narrator will tie in the theme, describe the action, explain the symbolism and give information on the New Orleans Mardi Gras in general. The presentation will follow dinner and after that a famous New Orleans orchestra will play for your dancing pleasure.

