EVALUATING TRAINING EFFECTIVENESS

DR. GENE S. MICHELI
Training Technology Department
Naval Training Device Center

How do we know when a training device is doing the job it was designed to do? It is only recently that we have begun to evaluate the effectiveness of training devices from the standpoint of transfer of training.

Many people are surprised to learn that while the Naval Training Device Center expends considerable time, money and effort on the analysis of training situations, development of training devices, utilization manuals, and research on training methods very little systematic effort has taken place on evaluating the effectiveness (or validity) of training systems from the viewpoint of its effect on on-the-job performance in an operational situation.

This state of affairs has been rectified by the Naval Training Device Center with the addition of a new division to the Human Factors Laboratory called the Training Effectiveness Division. Its mission consists of the following activities:

(1) Measuring the effectiveness of training systems and recommending improvements.

(2) Conducting field studies to measure the learning that takes place in teams and individuals undergoing training in NAVTRADEVCEN devices. (Measurements include degree and rate of learning, retention of learning, and the transfer of training (or carry-over) to operational situations.)

(3) Recommending changes in training systems of both design and patterns of use to improve training effectiveness.

(4) Analyzing cost benefits and measuring operational effectiveness as they affect training device design and use.

Since we have clearly stated objectives for training there is no need to carry out our training "blind." There is little justification for a training system which is not known to be effective. Mere opinions are not good enough, since these are often favorably biased. In any training situation we may reasonably ask whether the training system has resulted in any improvement of skill, or in making it easier or quicker to acquire the skill in the operational situation. In this we are in a similar position to the trainee. In order to make progress, the trainee must know what he is trying to do and how well he has succeeded. For example, in learning to fire a rifle there is little value in shooting at a target if the trainee cannot tell where the rounds hit. His skill will not improve much unless he receives feedback on where the bullets went.

So it is with the trainer. There is relatively little value in using a training system without trying to evaluate its effectiveness. With information on all parts of the training program, it is possible to decide how to make appropriate modifications in either trainer design, utilization patterns or curriculum.

Training devices are receiving ever greater acceptance. Their use is widespread and they are being assigned ever larger parts of the total training job. Let there be no question that training devices are effective and useful.

This paper suggests, however, that there has been little effort to evaluate the effectiveness of training done by a training device system.

The outcome of training is a function of both the capability of the training device (based on its design features) and of the effectiveness of the training offered on the device, that is, how the device is used.

There are several purposes to evaluation: (1) Improvement of a training program, either through changes in design or use, or both, regardless of whether the evaluation shows the training to be effective or not; (2) To measure the amount of transfer, that is, the extent to which the skills learned using the training device carry over to the operational situation. (This may be measured by the number of operational training hours saved or the level of performance in the operational situation.)

The latter information is needed to obtain answers to the questions of (a) How much synthetic training may be substituted for operational training consistent with safety, operational effectiveness, and readiness? and (b) What are the cost benefits of using trainers? Given the amount of transfer and the relationship of trainee proficiency to operational performance norms, some answers to these questions are possible. A comparison of trainee performance with operational norms will show the extent of progress toward the operational norm. Then the relative equivalence of training versus operational performance curves can be inferred. The cost of training in the operational situation can be compared with the cost in the trainer as a function of the specific performance levels reached for various stages of training.

What kinds of questions should be asked about a training device in terms of its training usefulness? First: Do trainees learn on the trainer? If a curve plotting performance on some parameter shows improvement over time, then learning occurred. Second: Do the trainees retain what was learned on the trainer? By having different trainees return to the trainer at varying intervals away from it and measuring before and after performance, a curve of forgetting can be plotted. Third: Is the training device providing the training for which it was intended, namely, to improve on-the-job performance in an operational situation, or at least, to shorten the time it takes for trainees to "get up to speed" on-the-job.

The third question, which refers to the concept of transfer of training, is the really crucial and ultimate question! Unless time and effort spent in the trainer have an influence on performance in the operational situation, training was not effective. It is even possible that negative transfer will make the trainee less effective in the operational situation than if he had not experienced the training at all (for example, if the trainee learned to attend to cues that either are not available or are the wrong ones to attend to in the operational situation). Generally, however, the skills acquired in the training situation are transferred, to some degree, to the operational situation.

Evaluation is conducted at one or more of several levels. Each level provides successively more information about the training situation. The first level is that of descriptive analysis. This is a minimum step that should be performed even when much more is possible. This first level of evaluation attempts to validate the content and procedures of instruction. A questionnaire is developed to determine answers about items such as the following: (a) Whether the content is relevant to the operational situation; (b) What the training objectives are; (c) The degree of realism of the trainee stations; (d) The monitoring, control and measurement features of the instructor's console; (e) The curriculum, guidelines or scenarios for problem presentation. Such data are collected by reviewing utilization manuals and scripts, by interviewing trainees and instructors, and by observing training sessions.

These data about the device and its utilization can lead to inferences about the effectiveness of the training system. The rationale behind this is that we know something about the effectiveness of the training when we know whether or not there are specified training objectives, proper monitoring, exercise control and feedback and adaption of training based on measurement of performance. An end product of this level of evaluation would be statements as to the positive features and deficiencies in design and utilization with recommendations for improvement where needed. However, the data collected on this basis are based on judgment and on previously developed criteria, so the conclusions that could be made are limited.

Even at this level, though, information on use patterns (or how the trainer is utilized in training) can be determined on an objective basis and can be very informative. For example, we can determine whether an attempt was made to train for skills and knowledges that are related to the skills and knowledges required in the operational situation, the extent to which the behaviors required on-the-job are accounted for in the training situation, or how the instruction is measured to determine whether the tasks being trained attain their objectives in terms of some criterion. On this last point, we do not expect instructors to have solutions to the criterion and measurement problems which research psychologists have not yet solved; we merely look for the extent to which some objective measures are taken and how they are utilized.

The first level (descriptive analysis) has the limitation of not giving any indication of whether the trainees actually learned anything. It results only in an inference about the degree of effectiveness. To demonstrate whether learning actually occurs requires measurement of performance.

Measurement requires another level of evaluation, namely, quantitative evaluation. The simplest test has been called non-comparative measurement, because it measures what occurs only in the current training situation. No comparison with alternate methods of training or with the operational situation is made. It obtains pretest and post-test scores. The difference in scores indicates the amount of learning that has occurred. The effectiveness of the training, however, is a function of the extent to which the content of the training can be judged to be related to the behaviors that are required in the operational situation. If the content can be judged relevant, then improved performance can be a reflection of training that would be effective on-the-job.

The measures of performance that might be used to indicate changes in trainee proficiency are: length of time or number of exercises to reach a specified level of proficiency; number (or types) of errors before reaching a specified proficiency level; final difficulty level reached; or pretraining and post-training comparisons.

Each of the two previous levels of evaluation can be done on a "not-to-interfere-with-training-basis." All that is needed is to observe training in process or to record measurements of proficiency. However if we want to compare the training with alternate methods of training or with performance in the operational situation in order to obtain data based not on logical bases but instead on actual comparisons, difficulties arise. For example, for a comparison of two methods (such as using or not using measured training performance or the use of a series of exercises graded in difficulty) control must be exercised. This might mean inserting standard exercises at the beginning, during, and at the end of training. As the alternate methods proposed may require instructor involvement, considerable cooperation will be needed. Such details have to be worked out with the staff of each trainer.

The final level of evaluation, is, as mentioned earlier, the crucial and ultimate one. This is the transfer of training test. It is the objective of training evaluation to determine whether training devices train the skills that they claim to train, and the extent to which those skills are trained. Training may be considered effective to the extent that on-the-job performance in the operational situation is improved.

To obtain information on the extent to which skills learned in the trainer transfer to the operational situation, the well-known transfer of training experiment is carried out. The basic research design is as follows (in which the performance in the operational situation of a group of trainees who received training on the trainer is compared with a group of trainees who received no training on the training device):

Group	Original Learning	Transfer Task
Training	Learn Task A	Perform or learn Task B
Control or operational	----------	Perform or learn Task B

The measurement of transfer of training may be a comparison of the training (or experimental) group with the operational (or control) group on the levels of performance shown in the operational situation, or the measure may be a comparison of levels of learning achieved during learning Task B. As you know, three kinds of transfer effects may be found: positive transfer, negative transfer or no effect. Both the kind and degree of transfer effect must be determined to evaluate training effectiveness.

The transfer of training experiments must be designed specifically for each training situation, because of the uniqueness of such factors as: selection of measures of performance in the operational situation which are compared to the measures of performance in the trainer to assess transfer of training; problems of reliability and validity of the measures; selection of experimental design and statistical analysis; and selection of control procedures.

Evaluation of training systems, then, is concerned with human learning and transfer of training. The problem of transfer extends beyond a training device. It is necessary to compare trainee performance on the equipment with later performance, hopefully in the operational situation. This comparison is more likely to be possible when the evaluators reach prior agreement with the users regarding the performance measurements to be taken and the comparison to be made.

We will now give you a brief description of selected examples of current and future projects of the Training Effectiveness Division of the Human Factors Laboratory.

An evaluation recently completed is that of the skill retention of those trained on the ASROC trainer. In this project performance changes by members of ASROC teams undergoing training at Norfolk were measured and their skills re-evaluated at periods ranging from eight to 32 weeks after training. Two rather straightforward conclusions were reached. One is that the trainees do in fact learn in the ASROC trainer. The other is that they rapidly forget what they have learned when they go to sea. It was concluded that shorebased team training should be made a regular part of the operating schedule of ASROC-equipped ships. The consensus expressed was that Device X14A2 practice was as good or better than at-sea practice, since it allows for multi-unit problems and unexpected contingencies. Despite these findings only four of 12 teams with convenient access to the training device had utilized it more than once during the year preceding this study. Problem exercises developed for this study are available and could be included in any Device 14A2 training program. We are following up on this.

A project on training system evaluation guidelines is concerned with the general problems of methodology when one attempts to assess trainee performance in the field and with providing guidance in solving these general problems. This study is developing criteria for determining how training devices can be evaluated and used most effectively.

An evaluation of an aircraft carrier landing trainer will determine the extent to which the skills learned in the trainer transfer to actual carrier landings. A transfer of training experiment will be conducted to compare pilots who are trained on the trainer and then proceed to flight training with pilots who do not receive practice on the trainer.

An evaluation of a tactical ECM trainer is also underway. One of its objectives is to determine the degree and rate of learning when using the trainer. This is to be accomplished through the use of standard exercises at the beginning, during, and at the end of training. We will also establish guidelines for the effective utilization of the trainer. It is hoped that we may be able to determine the extent to which the skills learned in the trainer transfer to operational situations.

Training effectiveness evaluations are also being planned for a joystick control submarine trainer, and for a weapons system trainer. In addition, plans are in progress for research on pilot training, non-flight officer training, and aircraft flight and tactics trainers to measure the effectiveness of such training.

Other projects in this program will give us information on how to better evaluate trainers or how to improve training. A task quantification study is one. Without quantitative information relating human performance and task characteristics, such things as the instructor's performance level and the trainee's learning rate in a new trainer are difficult to estimate until that trainer is operative. This research will develop quantitative indices of the characteristics of instructors' and trainees' tasks so that the effectiveness of a given amount and type of training on a given task can be predicted.

Since full and effective utilization of training devices requires acceptance by training personnel and trainees, a study will be conducted to identify the factors which lead to such acceptance. In addition, we will try to determine the extent to which the factors that are associated with acceptance of training devices are present in training devices which are designated as "effective" or "ineffective."

In summary, the Naval Training Device Center is taking part in an accelerated effort to develop measures of training system effectiveness. Ideally, both the designers of training systems and the users will come to welcome this and realize that long-term routine operation of training devices, in conjunction with formal evaluations, should result in modifications to the equipment and in its utilization. This realization, and its implementation, can tremendously improve the quality of a training system.