# QUANTITATIVE TASK ANALYSIS AND THE PREDICTION OF TRAINING DEVICE EFFECTIVENESS

MR. G. R. WHEATON AND DR. A. MIRABELLA
American Institutes for Research

Because of the enormous costs involved in the design and development of a complex training device, one can ill afford to adopt a "wait-and-see" attitude about the effectiveness of training which it provides. The primary problem confronting individuals responsible for military training, therefore, is how to plan for, design, and develop a training device from the very start, which will prove to be effective for a particular set of training objectives. But, given the requirements for training, how can one forecast or estimate how effective any specific design will be? For example, as designed will the device facilitate or inhibit ease of instructor operation (i.e., presentation of problem materials, monitoring and evaluation of student performance, provision of feedback)? Similarly, from the student point of view, will the design lead to rapid acquisition of skills and their positive transfer to the operational setting?

If answers to these types of questions could be given early in the design and development process, then a basis would exist for comparing and contrasting the "relative goodness" of alternative designs prior to actual development. By systematically evaluating alternative designs in terms of predicted training effectiveness, we might better be able to insure that the design finally adopted is better than other designs which may have been under consideration. This, after all, is the basic point. It would be desirable to have some early indication that the adopted design will provide superior training, relative to competing designs.

In the 25 years since World War II, few other training problems have received as much attention. The problem has come under repeated attack and has been approached from a number of different theoretical positions. Various methods have been conceived to help determine what should be trained and how training should be accomplished. Many of these approaches have shared the assumption that operational tasks possess certain critical characteristics which have specific implications for the design and utilization of training devices. It was hoped that this information, together with estimates of cost, would lead to training decisions which insured maximum returns for each training dollar invested. In spite of several efforts in this direction, however, the problem of prescribing the design of a training device, or of predicting its effectiveness, remains unsolved.

## THE EARLY YEARS--INTUITING THE SOLUTION

Historically, gross inadequacies in the design of training devices were often eliminated on the basis of shrewd guesswork. In the earliest approaches, design decisions were made by subject-matter specialists, who drew on experience and common sense, to solve training design problems. As a result, they often were able to make fairly sound decisions about the design of training aids and equipment, student and instructor stations, and other aspects of the training situation which might facilitate the learning experience. However, these early practitioners were artisans. Because of their experience, they were able to translate certain types of information about the job to be performed into requirements for training. As is true of all artists, however, they differed in terms of their conceptualization of and their approach to the training problems which faced them. As a result, some were highly successful in making sound training decisions. Others were not. Furthermore, because of the informal and implicit nature of their methods, it was difficult to train others in their use. But the major disadvantage of this approach lay in the difficulty of evaluating the proposed training solution prior to its adoption. Predictions as to the effectiveness of training were scarcely better than opinion.

## QUALITATIVE TASK ANALYSIS--A CERTAIN AMOUNT OF NAME CALLING

Because of the difficulties inherent in these individualistic methods, attention was focused upon the development of more formal and programmatic approaches. The results of these efforts were a number of job descriptive and task analytic procedures. Using these approaches, it became possible to describe jobs in terms of their major task components, and then to describe these components in terms of underlying task elements and activities. Description proceeded systematically through several levels. The earliest of these procedures (e.g., Miller, 1953; Miller & Van Cott, 1955) were designed to help specify those aspects of an operational task which should be considered as basic items of content in a training program. More recent efforts (e.g., Chenzoff & Folley, 1965), while retaining an interest in specifying the appropriate content of training, have also attempted to prescribe the manner in which training should be accomplished. Among the more advanced of these techniques is the Training Analysis Procedure (TAP) currently in use at the Naval Training Device Center. As described in the Fourth Annual Conference (Middleton, 1969), TAP is designed to aid in developing device requirements and translating these into functional characteristics of training hardware.

Other investigators have attempted to formalize training decisions by developing task classification systems having implications for training. These taxonomists shared the belief that basically different types of tasks did indeed exist. Given this premise, a logical step was to collect, sort, and catalogue tasks, casting them into their appropriate classes or families. For each identifiable class of tasks there might exist an unique or optimally effective set of training procedures. As a consequence of this thinking there have been several attempts to classify tasks and to specify for each class those training techniques which seem most appropriate (e.g., Willis & Peterson, 1961; Stolurow, 1964; Miller, 1969).

Many of the analytic and taxonomic methodologies developed to date have had their own particular problems. Most, however, have shared one weakness. They have become too dependent upon a process of name calling or labelling. A certain amount of name calling, of course, is inevitable, and in and of itself is harmless enough. But when a specific training decision may rest upon the label given a particular task element, and when we often can't agree on the label to apply (e.g., decision making, problem solving, inductive reasoning), then there is cause for concern. At this point the objectivity and reliability with which tasks can be described and analyzed are in doubt.

Such is the case with many of the classical task-analytic approaches. They have provided for the description of tasks in behavioral or functional terms (e.g., the behavioral taxonomy of Berliner, Angell, & Shearer, 1964; the functional descriptors employed by Gagne, 1962; and by Miller, 1966). These terms have been found difficult to apply unambiguously. This difficulty, coupled with the fact that the terms in use permit only qualitative distinctions to be drawn among different tasks, limits the utility of these approaches. While they may help to determine the basic content of training, and although they may even provide general guidelines about how training should be conducted, these approaches will not aid in the prediction of training effectiveness.

## QUANTITATIVE TASK ANALYSIS--PIGEON-HOLDING BY THE NUMBERS

In order to augment these conventional task-analytic procedures, therefore, the Naval Training Device Center has been seeking alternative ways of using information about the features of a training device and about the characteristics of the tasks performed within that device. Desired is a more reliable and objective method of task description which might be used to forecast training effectiveness. Underlying this effort there have been two issues of primary importance. First, would measures of training effectiveness (i.e., rate of skill acquisition, level of transfer) vary in some predictable manner as features of a training device were manipulated? Unless there was a relationship between these two sets of variables, prediction of effectiveness would not be feasible. Second, and even more basically, would it be possible to describe the critical features of a device reliably and along a number of quantitative dimensions? Unless such description were possible, there would be no way of investigating the relationship of interest.

To resolve these issues a research program was initiated by the Naval Training Device Center and the American Institutes for Research (AIR) which had three objectives. These involved: (a) Exploring ways of reliably describing features of trainee or instructor stations in quantitative terms; (b) determining whether the quantitative descriptors could be used to describe fairly complex devices; and (c) developing statistical methods for relating the quantitative descriptors to variations in device effectiveness.

In pursuit of these objectives, a variety of computer-driven and tape-based sonar training devices were examined, in order to identify features, which could be quantified. As a result of these efforts, a number of quantitative task-descriptive indices were assembled. These indices, developed by AIR, represented critical dimensions of the stimulus, response, and procedural aspects of trainee and instructor stations. Critical dimensions were those which, if manipulated, would be expected to affect level of (instructor) proficiency, rate of skill acquisition, or degree of transfer. Included among the indices were a variety of rating scales relating to such dimensions as work load, precision of responses, and response rate. Other indices were based on metrics such as the Display Evaluation Index (DEI) developed by Siegel and his co-workers (Siegel, Miehle, & Federman, 1963) and the several panel layout metrics developed by Fowler and his associates (Fowler, Williams, Fowler, & Young, 1968). The DEI is basically a measure of the effectiveness with which information flows from displays, via the operator, to corresponding controls. The panel-layout indices represent the extent to which general human-engineering principles have been applied to the design of hardware.

Application of the indices to four trainee tasks; (i.e., set-up, detection, localization, and classification) as represented in a variety of different sonar training devices, was attempted. This exercise demonstrated that most if not all of the indices could be used reliably to scale the extent and manner in which the trainee tasks differ across devices. By extension, therefore, the indices can be used to describe reliably and quantitatively how competing designs of the same device differ.

Can such quantitative information be related to measures of training effectiveness? In the final analysis, there is no way of specifying a priori which indices will bear a relationship to measures of performance, learning, or transfer. In order to explore the relationship between variations in design parameters and training effectiveness criteria, additional research is required.

Such research has been undertaken by AIR for NAVTRADEVCEN and is currently in progress. Training-effectiveness data are being collected both in the field and in the laboratory to provide criteria for regression analysis. The laboratory data are being obtained with the aid of a synthetic "trainer" which is an abstraction of the many sonar devices examined earlier in our research for NAVTRADEVCEN. The field data are being obtained through questionnaires, given to instructors and to trainee personnel. These questionnaires are designed to elicit estimates of the effectiveness of specific sonar trainers under standardized conditions. Use of the questionnaire, as opposed to more explicit performance evaluation, is being employed because of the well-known difficulties of accessing field equipment directly. Based on these field and laboratory data, regression equations will be developed. Such equations, if successfully validated, would permit the effectiveness of a given design to be predicted from knowledge of index values descriptive of that design. That is, we could then "pigeon-hole" by the numbers.

The trend away from purely qualitative task analysis procedures toward more quantitative techniques will result in improved training solutions. Such techniques offer promise in assuring the effectiveness of a training device earlier in the design process. Until these techniques are formally developed, predictions of trainee or instructor proficiency on particular trainer configurations will continue to be as much art as science.

## REFERENCES

Berliner, D. C., Angell, D., & Shearer, J. Behaviors, measures, and instruments for performance evaluation in simulated environments. Paper delivered for a Symposium and Workshop on the Quantification of Human Performance, (1964). Albuqurque, New Mexico.

Chenzoff, A. P., & Folley, J. D., Jr. Guidelines for training situation analysis (TSA). Technical Report 1218-4 (1965). Naval Training Device Center, Orlando, Florida.

Fowler, R. L., Williams, W. E., Fowler, M. G., & Young, D. D. An investigation of the relationship between operator performance and operator panel layout for continuous tasks. AMRL-TR-68-170, (December, 1968). Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

Gagne, R. M. Human functions in systems. In R. M. Gagne (Ed.) Psychological principles in system development. New York: Holt, Rhinehart, & Winston, 1962.

Middleton, M. G. The necessity of performing proper training situation analysis by engineering personnel. In Proceedings of the Fourth Annual Naval Training Device Center and Industry Conference (H. H. Wolff, Chm.), 18-20 November 1969. Report No. NAVTRADEVCEN IH-173. Naval Training Device Center, Orlando, Florida.

Miller, E. E. The development of a response taxonomy. Professional Paper 32-69 (1969). Human Resources Research Organization, Alexandria, Virginia.

Miller, R. B. A method for man-machine task analysis. Technical Report 53-137, (1953). Wright-Patterson Air Force Base, Ohio.

Miller, R. B. Task taxonomy: Science or technology? Poughkeepsie, New York, IBM, 1966.

Miller, R. B., & Van Cott, H. P. The determination of knowledge content for complex man-machine jobs. (1955). American Institutes for Research, Pittsburgh, Pennsylvania.

Siegel, A. I., Miehle, W., & Federman, P. Information transfer in display-control systems, VII. Short computational methods for and validity of the DEI technique (Seventh Quarterly Report). (Applied Psychological Services, 1963). Wayne, Pa.

Stolurow, L. A taxonomy of learning task characteristics. AMRL-TD 12-64-2, Aerospace Medical Research Laboratories, (January 1964). Wright Patterson Air Force Base, Ohio.

Willis, M. P., & Peterson, R. O. Deriving training device implications from learning theory principles, Vol. I: Guidelines for training device design development and use. Technical Report 784-1, (1961). Naval Training Device Center, Port Washington, Long Island, New York.

## SESSION II

Tuesday, 15 February 1972

Chairman: Mr. George Derderian
Head, Physical Sciences Laboratory
Naval Training Device Center