EVALUATION OF THE EFFECTIVENESS OF TRAINING DEVICES: LITERATURE REVIEW AND PRELIMINARY MODEL

George R. Wheaton, Andrew M. Rose, Paul W. Fingerman, Arthur L. Korotkin, and Dennis H. Holding
American Institutes for Research

Angelo Mirabella, Work Unit Leader
U.S. Army Research Institute for the Behavioral and Social Sciences

UNIT TRAINING AND EVALUATION SYSTEMS TECHNICAL AREA

This document has been approved for public release and sale; its distribution is unlimited.

U. S. Army
Research Institute for the Behavioral and Social Sciences

April 1976
1. The reports listed on Inclosure 1 are approved for public release with unlimited distribution (50 numbered ARI Research Memorandums, 74-1 thru 76-30).

2. These are among the previously unrecorded ARI reports which you identified to us 22 June 1979 as not in your retrieval system. The accompanying box contains at least one copy of each report for your retention and reproduction.

HELEN S. PRICE
Research Publications Group
Army Research Institute

<table>
<thead>
<tr>
<th>Accession For</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTIS General</td>
</tr>
<tr>
<td>DDC TAB</td>
</tr>
<tr>
<td>Unknown LOC</td>
</tr>
<tr>
<td>Justification</td>
</tr>
</tbody>
</table>

By


dr

A

REPLACES DD FORM 94, WHICH IS OBSOLETE.
EVALUATION OF THE EFFECTIVENESS OF TRAINING DEVICES: LITERATURE REVIEW AND PRELIMINARY MODEL

George R. Wheaton, Andrew M. Rose, Paul W. Fingerman, Arthur L. Korotkin, Dennis H. Holding
American Institutes for Research

Angelo Mirabella, Work Unit Leader

Submitted by:
Frank J. Harris, Chief
Unit Training and Evaluation Systems Technical Area

Approved by:
Joseph Zeidner, Director
Organizations and Systems Research Laboratory

J. E. Uhlaner, Technical Director
U.S. Army Research Institute for the Behavioral and Social Sciences

Research Memorandums are informal reports on technical research problems. Limited distribution is made, primarily to personnel engaged in research for the Army Research Institute.

This document has been approved for public release and sale; its distribution is unlimited.
## EVALUATION OF THE EFFECTIVENESS OF TRAINING DEVICES: LITERATURE REVIEW AND PRELIMINARY MODEL

### CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Background</td>
<td>1</td>
</tr>
<tr>
<td>Purpose of the Report</td>
<td>3</td>
</tr>
<tr>
<td>PREVIOUS MODELS AND METHODS</td>
<td>5</td>
</tr>
<tr>
<td>R. B. Miller's method</td>
<td>9</td>
</tr>
<tr>
<td>Demaree's Method</td>
<td>10</td>
</tr>
<tr>
<td>Willis and Peterson Method</td>
<td>15</td>
</tr>
<tr>
<td>Training Situation Analysis</td>
<td>18</td>
</tr>
<tr>
<td>E. E. Miller's Method</td>
<td>22</td>
</tr>
<tr>
<td>Shettel's Method</td>
<td>24</td>
</tr>
<tr>
<td>Caro's Method</td>
<td>27</td>
</tr>
<tr>
<td>Quantitative Task Description</td>
<td>29</td>
</tr>
<tr>
<td>Altman's Method</td>
<td>30</td>
</tr>
<tr>
<td>Conclusions</td>
<td>32</td>
</tr>
<tr>
<td>THEORETICAL POSITIONS AND ISSUES</td>
<td>35</td>
</tr>
<tr>
<td>Similarity Theories of Transfer</td>
<td>35</td>
</tr>
<tr>
<td>Mediation Theories of Transfer</td>
<td>40</td>
</tr>
<tr>
<td>Theoretical Issues in Modeling</td>
<td>42</td>
</tr>
<tr>
<td>Summary</td>
<td>48</td>
</tr>
<tr>
<td>SUBSTANTIVE ISSUES</td>
<td>50</td>
</tr>
<tr>
<td>Overview</td>
<td>50</td>
</tr>
<tr>
<td>Organization of Data</td>
<td>52</td>
</tr>
<tr>
<td>Category I - Training Variables</td>
<td>60</td>
</tr>
<tr>
<td>Category II - Device Variables</td>
<td>65</td>
</tr>
<tr>
<td>Category III - Task Variables</td>
<td>68</td>
</tr>
<tr>
<td>Summary and Implications: Training Principles</td>
<td>71</td>
</tr>
<tr>
<td>PRELIMINARY MODEL</td>
<td>79</td>
</tr>
</tbody>
</table>
FIGURES (Continued)

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.</td>
<td>The Skaggs-Robinson hypothesis</td>
<td>37</td>
</tr>
<tr>
<td>7.</td>
<td>Osgood's transfer and retroaction surface</td>
<td>38</td>
</tr>
<tr>
<td>8.</td>
<td>An alternative to the Osgood surfaces</td>
<td>45</td>
</tr>
<tr>
<td>9.</td>
<td>Preliminary structural model</td>
<td>81</td>
</tr>
</tbody>
</table>
EVALUATION OF THE EFFECTIVENESS OF TRAINING DEVICES: LITERATURE REVIEW AND PRELIMINARY MODEL

INTRODUCTION

BACKGROUND

In satisfying the Army's broad range of training objectives, the practice of using operational hardware has been giving way rapidly to the use of training equipment. A number of reasons have been advanced for this shift in training philosophy. These include reduced military budgets with consequent reduced availability of actual hardware for training purposes, reduced availability of large-scale training areas and ranges, advent of technology permitting safe two-sided maneuvers under quasi-realistic conditions, and finally, growing concern for the ecological damage which can arise from mechanized field engagements. These reasons provide sufficient justification for reliance upon simulative devices. But, there are additional considerations having to do with training technology per se.

It has become axiomatic among educational/training specialists that the complex processes of learning are not necessarily best served by "hands on" experience with real equipment. Instead, these processes may be better served by the simulative device, since it, unlike the operational equipment, can be specifically designed and employed to optimize such instructional features as feedback, scenario freeze and playback, sequencing of training events (i.e., easy to difficult materials) and finally, measurement of student achievement.

Employment of these features is often incompatible with the normal characteristics of operational equipment. For example, frequent practice in the operation of field artillery is constrained by cost of ammunition, availability of ranges, and safety considerations. Measurement of marksmanship performance in open field maneuvers is difficult without the use of sophisticated hit-kill indicators. Similarly, in the command/control area, evaluation of decision-making performance is difficult without the capacity to replay scenarios and examine decisions made at various stages of offensive or defensive engagements.

With the advent of the system engineering approach to the design of instruction, the simulator has become potentially even more important. It lends itself to the system approach particularly well and in ways not
possible with operational equipment. The simulator, hypothetically, can be fine-tuned to meet system-wide requirements; e.g., if the training system demands varying amounts and kinds of feedback to meet individual differences in styles of learning, then the simulative device can be designed with this kind of flexibility in mind.

Thus, for a variety of reasons, a trend has been established towards replacing operational equipment with equipment simulators in order to develop and maintain the skills of personnel. These reasons are compelling. Equally compelling, however, are a number of countervailing factors which require sober consideration.

The first of these is the cost of developing and producing simulators. This cost may be considerably more than the actual equipment because of the inclusion of instructional features. Understandably, the more flexible the simulator is with respect to these features, the greater the expense. This becomes especially true where the simulator employs an auxiliary computer to control stimulus presentation and/or to record student performance.

The question of cost is confounded by a second factor. There is limited knowledge concerning the effectiveness of various instructional features, particularly as measured by transfer of training (TOT). The question is further confounded by lack of knowledge about how much of the variance in learning behavior is attributable to interactions among several potent variables within any given training medium. Little is known about how rate of learning and transfer of training are influenced by interactions among task characteristics, training device design, trainee attributes, and training techniques. Until the relationships among these components are thoroughly explored and documented, the problem of designing an effective training system, on other than a trial and error basis, will remain unresolved. In fact, the capacity for building training system components, including sophisticated training equipment, audio-visual devices, and classroom training aids, has far outstripped knowledge about how to design them, and how and when to use them vis a vis the specific behavioral objectives to be achieved.
Thus, while the need for increasing reliance on training devices is clear, it is equally evident that their cost-effectiveness cannot be taken for granted. Some means must be found for evaluating training equipment and for doing so within a broad systems context which includes other classes of variables which may significantly limit training device effectiveness. Ideally, this evaluation should be feasible during early stages of the device design and development cycle. In this manner alternative designs could be contrasted in terms of their predicted effectiveness, with the best design package being selected for prototype development and broad-scale procurement.

In order to conduct such an evaluation, a model or conceptual framework is needed which will provide systematic guidelines for predicting the effectiveness of a given training device at various stages in its development. The model needs to take into account what must be trained, who must be trained, and how the training is to be accomplished. The kind of model which is required is not yet available.

**PURPOSE OF THE REPORT**

The present report is the first in a series describing a program of research whose goal is the development and eventual validation of a method for predicting training device effectiveness. As the lead publication in the series, this report presents a preliminary model for the prediction of one of the most important aspects of training effectiveness—transfer of training.

In developing the preliminary model, every attempt has been made to examine and, if possible, to build upon previous efforts. Toward this end, several different kinds of literature potentially bearing on the prediction of device effectiveness have been exhaustively reviewed, reduced, and analyzed. Previous methods and models dealing with the design or evaluation of training programs were examined. General theories of transfer were studied as were the specific constructs believed to mediate transfer. Finally, a host of substantive issues were examined, particularly in terms of empirical data on specific variables and their impact on transfer. The report describes and discusses this information and, when
appropriate, indicates its incorporation into the model.

In conducting the review, over 2,000 abstracts were screened for possible relevance. Based upon this initial evaluation, over 500 documents were eventually acquired, more than half of which were directly relevant to either the structure of the model or to issues surrounding its application. To the best of our knowledge, only one previous review, compiled by Bernstein and Gonzalez (1971a, c) and indexed by Blaiwes and Regan (1970), has covered, to a comparable extent, what has proved to be a very diverse and fragmented literature.

In the following four sections of the report the results of the literature survey are described together with implications for a preliminary model for use in predicting training effectiveness. The following section discusses previous attempts to prescribe or to predict effective training. The subsequent section outlines the major psychological theories of TOT, together with their implications for predictive model. The next section summarizes the empirical literature on learning and transfer and paves the way for presentation of the AIR preliminary model. The model is described in the final section.
PREVIOUS MODELS AND METHODS

In generating the basic concepts which would underlie the initial model development, considerable attention was given to those models and methods already available and discussed in the literature. There are several such major systematic approaches and each will be described and critically discussed in the remainder of this section. It would appear that a good and generally applicable framework for comparing and contrast these systems is an information flow diagram. Therefore, each of the models will be discussed, wherever possible, in terms of its Purpose, Scope, Inputs, Processing and Output.

First, with regard to Purpose, most of the methods currently on hand were designed to analyze and prescribe training needs (thereby permitting determination of appropriate device design) rather than to predict device effectiveness. In the case of the prescriptive approaches, the goal is to design or specify an optimal training system, including the necessary devices and their manner of utilization. It is conceivable that such methods could be used to predict a device's effectiveness, but only at a gross level. For instance, one would derive the optimal device design and then would compare it to the training device under examination, concluding whether or not the latter was optimal. By the same token, those systems primarily concerned with prediction would, in addition, allow one to design optimally by manipulating parameters to simulate various potential devices until an optimal one is found. Thus, while both systems have utility, it is the existing predictive models which are most relevant to the present effort. The prescriptive models are of interest, however, to the extent that they suggest dimensions important for prediction which are not represented in current predictive methodologies.

The Scope of existing models refers to the variables upon which they focus. Current systems appear to vary with regard to their inclusion or emphasis on learning (acquisition) as distinct from the transfer of the results of that learning. Most existing models clearly emphasize one or the other type of measure. Furthermore, many of the transfer-oriented models admit to predicting potential transfer, stating that actual transfer will depend
on the use to which the device is put. Thus, while they infer that principles of learning effect the transfer, such principles are not included in the model itself.

The input factors include the type and level of the data required by the model as well as the feasibility of obtaining those data. For the most part, this issue revolves around the depth of the task analyses. In the existing models this varies from fairly general task descriptions to highly detailed micro-analysis. Detail of description or depth of analysis appears to be a key dimension since most of the models reviewed utilize some form of task description/analysis. They variously address the operational task to which training is directed, or both the training device task and the operational task. Some substitute a micro-analysis of stimuli and responses for more traditional task analysis.

The task-analytic models assume that if the analysis is appropriately performed, it will "permit the classification of tasks into sets or categories which are relatively homogeneous and invariant with respect to principles of learning, training techniques, etc." (Wheaton, 1968, p. 8). In other words, it is assumed that information about a given task category can be used to indicate the best method for training tasks within that category. Most of these models are presented in a matrix form to show task type versus training principle associations. The emphasis in these systems is clearly on acquisition, transfer being assumed to occur as long as the tasks trained (or knowledge learned, or skills acquired) in the synthetic training situation are the same as those in the operational situation (Miller, 1954a, b; Haggard, 1963). These systems also usually tend to focus on prescription rather than prediction, and empirical support for the relation between tasks and training techniques is often lacking. The general approach suggested by the task analysis systems is illustrated in Figure 1.

The micro-analytic models, as mentioned above, focus on a comparison of the stimuli and responses involved in the operational and training situations. The emphasis on comparative assessment suggests that these models may be either predictive or prescriptive. (See Figures 2a and 2b.)

Before proceeding to discuss existing methods in terms of the kinds of distinctions mentioned above, it should be noted that there is an
Figure 1. Prescriptive task-analytic approach (After Haggard, 1963).
Analyze Stimuli and Responses in Operational Task

Analyze Stimuli and Responses in Training Task

Compare (Consider Realism, Criticality, etc.)

Predict Transfer as a Function of Similarity

Figure 2a. Predictive micro-analytic approach (After Caro, 1970).

Analyze Stimuli and Responses in Operational Task

Create Training Task to be as Similar as Possible, with Possible Addition of Techniques to Facilitate Training (e.g., Augmented Feedback)

Figure 2b. Prescriptive micro-analytic approach (After Smode, 1972).
alternative to models altogether. Jeantheau (1971) and Jeantheau and Andersen (1966) have suggested that the appropriate way to deal with training device effectiveness is to measure transfer of training (TOT) rather than to predict it, since if one can measure TOT, the necessity of predicting it is obviated. While their discussions deal primarily with appropriate experimental designs and measures, the considerations they raise are important in the context that prediction should parallel the potential outcome of a measurement experiment. These methodological issues will be considered in a subsequent report in this series.

We now turn to a description of some of the methods previously proposed for dealing with training device effectiveness.

R. B. MILLER'S METHOD

Miller's method (1954a, b, 1960; Smith, 1965) is basically a task-analytic procedure, designed to derive from an analysis of the operational tasks several kinds of information bearing on training decisions. These decisions include: 1) functional training requirements, i.e., a description of the kinds of training needed; 2) a gross specification of the kinds of devices which would be appropriate; and 3) an indication of the way in which certain tasks can be grouped for training on a single device.

Miller begins by breaking down the operational situation into missions (if tasks vary from mission to mission), and within each mission, prepares a task-time chart which enumerates tasks, and groups them as a function of time (i.e., successiveness in the chronological cycle) and kind (i.e., similarity of skill or associated equipment operations). Further, a time diagram is prepared to show continuity among tasks as well as time-sharing considerations. The conditions under which each task is performed are then listed. On the one hand, this step aids in detecting subsidiary or contingent tasks such as dealing with enemy fire while aiming a missile. It also permits identification of those conditions which are likely to degrade performance. Finally, each task is subdivided into subtasks or "activities."

1He defines task as a set of activities related to each other by proximity in time and a common purpose.
Given these inputs, a table is then prepared which lists tasks and subtasks across the top, and training strategies along the side. The latter, as shown in Table 1, consist of compound statements about stages or kinds of training and associated trainer types. All of the subtasks to be trained are then evaluated against the training strategies, on the basis of which appropriate stages of training and trainer types are chosen. The matching of content with types of training is presumably accomplished on the basis of the analyst's expertise. Next, the charted subtasks are grouped to indicate those tasks and training phases which can be trained together on a single device. Finally, the specific hardware needed to implement the selected training strategies is identified. This final step of actually prescribing training device design is accomplished following guidelines (Miller, 1960) as to which devices are best for the various kinds of training.

Miller's model is clearly prescriptive rather than predictive in purpose and while its scope includes both learning (acquisition) and transfer criteria, it leans most heavily on the former. The input data are derived from a task analysis at the behavioral level. In essence, Miller's procedure describes a systematic way of recording information about training. However, the descriptors used are not adequately defined, nor are systematic procedures provided for combining the information in order to design devices or to predict their effectiveness. The major drawbacks to this approach lie in the ambiguity of the procedures required and a heavy dependence on unspecified analyst experience to provide the input. Smith (1965) comments, "Just how these conclusions [about devices] fit the [task-training type] matrix...seems unclear. ... The gap between [training] requirement and equipment seems to be bridged no less intuitively here than in less systematic development."

DEMAREE'S METHOD

While Demaree's method (Demaree, 1961; Smith, 1965) is conceptually related to other task-analytic approaches, it is more systematic than most. His system for analyzing operational situations to produce training recommendations is broken down into two major stages. The first stage gathers "Training Equipment Requirements Data," which are used to generate
<table>
<thead>
<tr>
<th>TYPES OF TRAINERS</th>
<th>STAGES OF TRAINING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Familiarization Trainers</td>
<td></td>
</tr>
<tr>
<td>1. Demonstrators</td>
<td>Demonstration of principles of operation, etc.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Nomenclature and Locations</td>
<td>Nomenclature and locations training</td>
</tr>
<tr>
<td>Instructed-Response Trainers</td>
<td></td>
</tr>
<tr>
<td>1. Detection of Conditions</td>
<td>Detection of conditions training</td>
</tr>
<tr>
<td>2. Identification of Conditions</td>
<td>Identification of conditions training</td>
</tr>
<tr>
<td>3. Problem Solving; Decision Making</td>
<td>Decision making; problem solving</td>
</tr>
<tr>
<td>4. Instructed-Response Procedures</td>
<td>Instructed-response procedures</td>
</tr>
<tr>
<td>Automatized Skill</td>
<td></td>
</tr>
<tr>
<td>1. Advanced Tracking</td>
<td>Advanced tracking training</td>
</tr>
<tr>
<td>2. Job Segment Trainer and Simulators</td>
<td>Automatized, coordinated response and work context training</td>
</tr>
</tbody>
</table>
the functional requirements for the training situation. The second stage gathers "Training Equipment Selection Data," which are designed to produce specific equipment recommendations to fulfill the functional requirements. This approach appears to be a forerunner of the Naval Training Device Center's Training Situation Analysis, developed by Chenzoff and Folley and discussed later in this section.

Demaree's Training Equipment Requirements Data stage begins with a task analysis not unlike R. B. Miller's. The operational task is analyzed into training functions which consist of statements of what the trainee needs to learn. There are four classes of training functions in his method, and each is related to an appropriate type of training device on an a priori basis. (See Figure 3.) Functions are distinguished from one another by searching for sets of associated activities where the skills and knowledge required comprise a "unitary training requirement" (i.e., a requirement for which a single piece of training equipment will be needed). He then subdivides each training function into discrete tasks and relates the degree of realism required in the appropriate device for each task or activity comprising the function. The final step in the task analysis is to rate for each task the performance proficiency criteria desired at the end of training on the device. This rating is performed on a seven-point scale which ranges from "no experience or training required" to "has complete understanding of the task; can do it completely and accurately without supervision; can apply the technique and skills to other equipment or situations."

Once the training functions have been identified and described in terms of appropriate devices, degrees of realism, and proficiency criteria, each is further considered in terms of a number of utilization factors. These include such considerations as:

1. the use of existing equipment;
2. the training context including the course of instruction, the sequence and mix of training, and the expected nature of materials to be used concurrently;
3. trainee characteristics;
<table>
<thead>
<tr>
<th>Training Functions</th>
<th>Simulator</th>
<th>Training Device</th>
<th>Training Aid</th>
<th>Training Part</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Learning of Knowledge</td>
<td>X</td>
<td>X</td>
<td>XX</td>
<td>X</td>
</tr>
<tr>
<td>2. Learning of Skills and</td>
<td>X</td>
<td>XX</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Task Components</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Learning Whole-task</td>
<td>XX</td>
<td>X</td>
<td>XX</td>
<td></td>
</tr>
<tr>
<td>Performances</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Learning Rated Task</td>
<td>XX</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Performances</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Best device noted by "XX"

Figure 3. Training equipment for various kinds of training functions (After Smith, 1965).
4. the amount and the nature of usage (individual or group practice, demonstration, amount of time per trainee or group, number and rate of trainees or groups); and

5. instructor requirements (availability and qualifications).

From these considerations, hours of individual use, group use, total use, and the number of devices required are estimated.

The final step in assembling the "Requirements Data" is to consider the impact of eleven "Training Effectiveness Characteristics" on the training of each specific task. Each is rated by the analyst to determine what "level of complexity" (usually represented by degree of fidelity) is required for each training function. The eleven characteristics are:

1. Equipment representation (fidelity),
2. Trainee responses (fidelity),
3. Trainee coverage (crew or individual training),
4. Trainee orientation (is orientation accomplished by the device? the instructor?),
5. Performance aids (augmentation/guidance),
6. Information feedback (fidelity),
7. Programming (sequence of problems, etc.),
8. Proficiency evaluation (trainee performance evaluation),
9. Effective use time (maintainability and reliability, relative to operational equipment),
10. Acceptability (to students and instructors), and
11. Time availability (undefined by Demaree).

Demaree provides specific rating criteria only for the first characteristic in the list. The rest are left up to the analyst to rate, although the scales for doing so are defined.

All of these data are then compiled in a matrix which serves essentially as a summary of the judgments and ratings made to this point, and shows the presumed interactions between task type (training function) and each of the other classes of information.
Since the second stage of Demaree's analysis, "Training Equipment Selection Data," is roughly parallel to Chenzoff and Folley's "Training Analysis Procedure," it will not be covered here. This method will be reviewed later in connection with their approach.

In summary, Demaree's approach to training device design is much like R. B. Miller's and Chenzoff and Folley's. Its purpose is prescriptive, and the scope is primarily directed at acquisition criteria (although the effectiveness characteristics are partially related to transfer). While Demaree apparently provides a more systematic process for gathering input data than Miller, especially with the use of well-defined rating scales, the entire procedure is still subjective and heavily dependent on expertise. Justification is provided neither for the training function-device type associations, nor for the training function-fidelity rating guidelines. Other scales are discussed, but the manner in which they are to be used is not described by Demaree. Unlike Miller, however, Demaree does include provisions for recording information on trainees and desired achievement levels. It should be noted that much of the information generated on utilization and effectiveness are never used in the analysis, having presumably been gathered for some later, unspecified, cost-effectiveness tradeoff analysis. Finally, there is little empirical support for the specific guidance and criteria offered.

WILLIS AND PETERSON METHOD

The method developed by M. P. Willis and R. O. Peterson (Willis and Peterson, 1961; Smith, 1965; U.S. Naval Training Device Center, 1972a) represents the next step in the conceptual development of a task-analysis based, acquisition-oriented scheme for the prescription of training. Willis and Peterson make explicit the key assumption of Miller and Demaree, and of most other task-analysis based schemes: the crucial item for analysis of training device effectiveness is the interaction between specific task taxonomic categories and training situation variables. They formally express this interaction in a matrix, one axis of which represents categories of tasks, while the other axis represents training "principles" or variables. The choice of specific design options in each of the training variable categories is assumed to be at least partially dependent on the
task category under examination. This kind of an interaction is only implicitly represented in the matrices employed in the Miller and Demaree systems.

Willis and Peterson began by conducting an extensive review of the learning theory literature, on the basis of which they derived thirteen basic principles of skill acquisition. They then developed a detailed, nineteen-category behavioral taxonomy. The applicability of each principle to each task category was then considered. This step led to development of a nineteen by thirteen matrix, each cell of which presumably could provide particular design guidelines for specific kinds of tasks. (See Figure 4.) It should be noted here that not all cells of the matrix are necessarily unique in terms of different guidelines, since some principles are applicable to more than one behavioral task category. Further, information is not available for all of the cells, since occasionally no particular training guideline could be extracted on the basis of existing knowledge for the specified task-principle interaction cell.

To use the system, one begins by analyzing the task. A task description is recorded, a list of critical activities is compiled, and finally, the activities are classified according to the nineteen available task-behavior categories. (Note that this system does not involve the specification of unitary training requirements, or the time relationships among activities, as do the systems of Miller and Demaree; nor is provision made for direct use of information about time-sharing or crew interdependence.) Once the task to be trained has been described in terms of the nineteen behavioral categories, one final step remains. In theory, at least, the training design guidelines associated with each category are identified and organized into an overall training strategy. As pointed out by Willis and Peterson (1961, p. 1), accomplishing this last step is not always easy.

The advantage of this system is that it seems to require less expertise than is needed in the two preceding methods. If the analyst can assign the task categories appropriately, the task-principle matrix supplies the guidelines. These represent the functional requirements for the device. However, Willis and Peterson admit that the task categories are intuitively rather than empirically derived. Thus, significant expertise may be required
**Figure 4. Task by principle matrix (Willis and Peterson, 1961a).**
(to be consistent with their intuition), and, further, some other specific categories may be more appropriate to the task of organizing the learning principles. Similarly, a good deal of subjectivity surrounds the guidelines which are to be derived from the task-principle matrix. The guidelines, at least partially based on the ingenuity and intuition of the authors, lack formal empirical support. The various principles were derived exclusively from theories of learning, there having been little consideration of principles of transfer.

In summary, Willis and Peterson's approach is prescriptive in purpose as are the two previously described systems. The scope of the approach is still primarily acquisition-oriented, and the basic input data are derived via task analysis. Further, despite the apparent sophistication of this model, it is in reality nearly as subjective as the R. B. Miller and Demaree systems. It does, however, represent a step forward inasmuch as the assumptions about each interaction are explicitly stated, and formal use of learning theory is made in deriving the system.

TRAINING SITUATION ANALYSIS

As described by Chenzoff and Folley, this approach actually consists of two methods: 1) Task Analysis Method (TAM) developed by Chenzoff and Folley (Chenzoff, 1964; Folley, 1964; Chenzoff and Folley, 1965); and 2) Training Analysis Procedure (TAP) developed by Van Alberti, Jeantheau, Gorby, and Parrish (1964), and revised by Chenzoff and Folley (1965).

TAM is a multi-stage system for task analysis, each stage of which essentially involves a finer and more detailed level of analysis, coupled with a focusing in on problems relevant to the training analyst. This last consideration is particularly important since it helps insure that only relevant details will be dealt with as the analysis proceeds. TAM consists of five separate stages as described below.

The first stage of TAM consists of familiarizing oneself with the system under study. This is accomplished by developing a System Block Analysis (SBA), which represents a flow chart of the major blocks of tasks or system operations, all of which are directed toward the system goal (e.g., hitting a target). The time sequence of system operations is indicated
where possible, contingency branches are noted, and equipment used in each block is identified. Each block designated in the SBA serves as input to the second stage of TAM. In the second stage, Task-Time Charts (TTC) are prepared which identify each task in the block, the appropriate operator or position, the typical time for completing events in the block, the typical time for each task, the coordination requirements among tasks and positions, and any adverse conditions which would affect tasks being performed during a particular time segment.

In the third stage of analysis, each task identified earlier is subjected to further scrutiny in a Functional Task Description (FTD). As the first step in FTD, the maximum permissible time to complete each task is determined, as well as the typical completion time. Next, each task is analyzed into seven functional categories or activities (see Figure 5), and estimates are given of the amount of time spent in each activity as well as the amount of attention required. Contingencies that may disturb normal performance are identified, and appropriate branches are listed. Finally, adverse conditions specific to the task are listed, and their probability of occurrence and severity are estimated. The fourth stage of analysis, Behavioral Details Description (BDD), primarily entails a second look at the FTD in terms of "psychological characteristics." This step is designed to provide more detailed information about the various functional categories or activities comprising each task.

According to Chenzoff and Folley, the first four stages of TAM are designed to provide a task description, which is then to be translated into a set of Functional Training Requirements (FTR). The procedure for accomplishing this fifth and last stage of TAM is unclear, as the authors indicate.

"However, an explicit, step-by-step procedure has not as yet been devised for the FTR stage . . . A person who is fairly sophisticated in both the methodology of TAM and in the planning of training programs should be able to use the TAM task descriptions to derive functional training requirements for a system. Until the process whereby these decisions are made is stated explicitly, however, the training solution adopted for any given system will be partially based upon the judgment of the 'expert' making these decisions (Chenzoff and Folley, 1965, p. 73)."
Figure 5. Functional task description (adapted from Chenzoff & Folley, 1965).
This is indeed too bad, since TAM is by far the most systematic and formally
developed method encountered so far. The first four stages are supported
by detailed definitions and examples, so that good reliability of description
would be expected. Moreover, the data generated during stages two, three,
and four seem particularly appropriate as input to an FTR stage, on the
basis of which the training solution would be formulated.

In any event, if one assumes the existence of an FTR stage, the next
step in the overall method is Training Analysis Procedure (TAP). The goal
of this analysis is to rank tasks in terms of the greatest training benefit
which is anticipated per dollar expended. The rating is based on the
ratio of expected improvement in performance to the dollar cost of that
improvement. Expected improvement in performance is represented by a Figure
of Merit (FOM), and is obtained by estimating speed and accuracy in each
task for both untrained and trained operators. FOM is then defined as the
"percentage improvement in system performance as a result of training on
individual tasks" (p. 79). The calculations involved in estimating FOM
for each task depend on various task characteristics (e.g., repetitive or
non-repetitive, amount of monitoring, rate task or fixed-sequence, bottleneck
or non-bottleneck, etc.).

Once an estimate is available a FOM/cost ratio is calculated for each
task being considered for training, where the largest ratio represents the
most system improvement per dollar. This index is then used to select
tasks for training.

As with TAM, the TAP data collection and recording procedures are laid
out in exceptional detail. Unfortunately, however, there are major problems
associated with TAP. In a study by Bertin, Colvin, Benfari, Lanchony, Logan,
Metlay, Suwara, and Wallach (1963), TAP was tried out for two systems. The
major problem was in obtaining reliable untrained performance time and
accuracy estimates, to which the model is very sensitive. Such estimates
were unobtainable in any practical fashion. In addition, application of
TAP to other than simple systems proved impractical due to ambiguity
in some of the required decisions. Finally, it was suggested that TAP was
insensitive to task criticality, and did not address the amount of training
time required (as opposed to cost), the use of part-task training, or
the relations of some part-tasks to system performance.

Fundamentally, the system is similar to others previously described in that it is prescriptive, is focused on acquisition rather than transfer, is dependent upon a task analysis, and has little formal empirical basis. Unlike the others, however, it is highly systematic, may require the gathering of less task data (often based on interviews rather than extensive observation), and provides a gain in efficiency without adverse effects on the quality of the data. The TAM component represents the distillation of other methods and is intuitively compelling despite the lack of adequate empirical support. The TAP component is less well developed. If one leaves aside TAP, however, the remainder of TSA (TAM, and an FTR analysis which remains to be developed) looks extremely promising. The approach appears to pinpoint critical information needed by the training analyst, and requires less detailed raw task description than some other methods (Smith, 1965).

E. E. MILLER'S METHOD

Structurally, E. E. Miller's (1966) method of analyzing tasks for training requirements is generically similar to that developed by Willis and Peterson. His system is based on a matrix which crosses task taxonomic categories with classes of training strategies, and is derived from a review of the literature, tempered with a certain amount of intuition. One uses his matrix by finding the task of interest, and reading training guidelines which he has generated for each indicated cell. (See Table 2.) His four task categories shown across the top of the matrix include: 1) reactive, adjustive (e.g., tracking); 2) reactive, choice (selection from a set of responses as a function of specifying stimulus cues, e.g., discrimination); 3) developmental, procedural (e.g., starting a car, or other sequential military type procedures); and 4) developmental, evolutionary (tasks such as hitting a baseball, which while procedural, requires the development of fine skill or techniques as opposed to remembering a list of steps). Down the side of his matrix are classes of training techniques and acquisition variables which may be manipulated for impact on training. Numbers within the matrix cells refer to training guidelines. As presented in his report (1966), for example, number 1 is:
Table 2

INTERACTIONS OF TASK CATEGORIES WITH TRAINING STRATEGIES

<table>
<thead>
<tr>
<th></th>
<th>Reactive</th>
<th>Developmental</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adjust-</td>
<td>Proce-</td>
</tr>
<tr>
<td></td>
<td>ive</td>
<td>dural</td>
</tr>
<tr>
<td>A. Operational Conditions of Practice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Representation of task environment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Unmodified</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Modified</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Stimulus predifferentiation</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(2) Response practice under progressively more</td>
<td></td>
<td></td>
</tr>
<tr>
<td>difficult conditions</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>2. Analysis into subtasks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Performance requirements information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Supplementary knowledge of results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Incentive manipulations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Progress Diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Utilizing knowledge of results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Clarify goal state</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Call attention to benchmarks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Supplementary (early) knowledge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Process conception</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Response set for effective feedback</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Movement consistency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Avoid responses which mask feedback</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Overt response patterns</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Sensitivity to cue indicating moment for response</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Response anticipation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(From Miller, 1966)

23
"Stimulus predifferentiation methods may apply to adjustive responses when the cue function is unclear, or when there are no reference marks. In such cases, the function of stimulus predifferentiation is to clarify the feedback quantity for the subject."

Guideline number 7 states:

"Because developmental, procedural tasks are matters of remembering the response rather than of performing skillfully, they provide an ideal situation for modifying response demands in speed, force, or amplitude, or form. Also, the physical situation for such procedural tasks generally allows for modifying both the time and the form of the response."

Miller's guidelines, although couched in more psychological language than those of Willis and Peterson, serve the same function of helping the analyst to specify the functional requirements for a training device. His approach is generically similar to Willis and Peterson's in that it is prescriptive, acquisition-oriented, and still based on task-analytic data for inputs. There are no explicit procedures described for processing the task analysis inputs other than examples. There is little or no empirical support for the resulting recommendations.

SHETTEL'S METHOD

A variation on the prescriptive scheme has been developed by Shettel (Shettel and Horner, 1972a, b). This approach, entitled "Training Event Analysis" (TEA), is designed to translate task-analytic data into training device requirements. The method consists of identifying significant learning elements, categorizing these with respect to type of behavior, identifying the training techniques related to each behavioral category, and arranging these techniques into a training program.

As the first step in the derivation of training requirements, a detailed task analysis is conducted. During this effort emphasis is placed on a level of description which is both necessary and sufficient (i.e., neither too general nor too microscopic) to describe the major behavioral events involved in the task. Consequently, description may be at the sub-task, task-element, or step level. Given these data, a Training Overview
is prepared for each task, including: 1) the task title; 2) a definition of the task, including the initiating and ending events, what is not included, and prerequisite skills, if any; and 3) a list of significant learning elements, i.e., any element which is difficult to learn, or which is critical and would prevent completion of the task if not learned.

A fourth and very important component of the Task Training Overview consists of a statement about training rationale. In this statement the significant learning elements are categorized with respect to four different "types" of behavior (1. perception/monitoring; 2. procedure following/execution/communication; 3. decision making/mediation; and 4. motor/execution). The training strategies and techniques deemed most appropriate for each type of behavior are determined. This is done by referring to a matrix in which a set of training techniques is rank-ordered for appropriateness within each behavior category. (See Table 3.) This information is then synthesized into a statement describing the methods and approach for combining the Significant Learning Elements into a cohesive set of Training Events. Finally, each Training Event is listed and then further defined by a precise description of what the student should learn, the necessary stimuli for instruction, a description of the responses the student will make, the feedback required, and the recommended training techniques.

The next step is to determine the Functional Training Requirements for each Training Event in each task. Two kinds of requirements are defined: those needed to implement the recommended training technique, and those physically required to perform the task. The Functional Requirements are described and organized into categories, including:

1. Student station requirements (necessary stimuli or information requirements; necessary controls; necessary feedback)
2. Visual display requirements (stimuli as required in the Training Event specification)
3. Motion requirements (as needed to implement the Training Event)
4. Sound requirements (as correlated with stimuli and responses)
Table 3
TRAINING TECHNIQUES RELATED TO CLASSES OF BEHAVIOR

<table>
<thead>
<tr>
<th>A. Perception/Monitoring</th>
<th>C. Decision Making/Mediation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variation of examples</td>
<td>Variation of tolerances</td>
</tr>
<tr>
<td>Variable-order presentation</td>
<td>Variation of examples</td>
</tr>
<tr>
<td>Variation of tolerances</td>
<td>Variable-order presentation</td>
</tr>
<tr>
<td>Distributed trials</td>
<td>Repetition</td>
</tr>
<tr>
<td>Repetition</td>
<td>Distributed trials</td>
</tr>
<tr>
<td>Prompting</td>
<td>Fixed-sequence presentation</td>
</tr>
<tr>
<td>Fixed-sequence presentation</td>
<td>Prompting</td>
</tr>
<tr>
<td>Vanishing</td>
<td>Vanishing</td>
</tr>
<tr>
<td>Massed trials</td>
<td>Variation of pace</td>
</tr>
<tr>
<td>Variation of pace</td>
<td>Massed trials</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. Procedure Following/Communication</th>
<th>D. Motor/Execution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed-sequence presentation</td>
<td>Distributed trials</td>
</tr>
<tr>
<td>Repetition</td>
<td>Variation of tolerances</td>
</tr>
<tr>
<td>Variation of examples</td>
<td>Fixed-sequence presentation</td>
</tr>
<tr>
<td>Prompting</td>
<td>Repetition</td>
</tr>
<tr>
<td>Vanishing</td>
<td>Variation of examples</td>
</tr>
<tr>
<td>Distributed trials</td>
<td>Prompting</td>
</tr>
<tr>
<td>Massed trials</td>
<td>Variation of pace</td>
</tr>
<tr>
<td>Variation of tolerances</td>
<td>Vanishing</td>
</tr>
<tr>
<td>Variation of pace</td>
<td>Variable-order presentation</td>
</tr>
<tr>
<td>Variable-order presentation</td>
<td>Massed trials</td>
</tr>
</tbody>
</table>
5. Instructor station requirements (necessary control of stimulus presentation, and ability to supply feedback)

6. Interactive requirements (between controls and visual displays when tracking, etc., plus interaction between student and instructor)

At this point, Shettel's system has specified the optimal requirements for training. Unlike most other models, however, he goes beyond this point, and develops a system for evaluating real devices relative to this optimum, and for predicting the relative effectiveness of two or more devices. Using a fairly complicated rating procedure the analyst can obtain an estimate of the effectiveness of a device for specific training events relevant to a hypothetical optimal device. Two devices can be contrasted by comparing their two total effectiveness capability scores.

Shettel's model is still prescriptive in purpose and acquisition-oriented with regard to scope. Inputs come from both task analysis and rate judgments, at a level intended to be "neither too specific nor too general." Most of the procedures for processing the data tend to be explicit. There has been a serious attempt to apply the model systematically and thus establish some empirical basis for it. However, the criticisms posed against other task-analytic based models apply to this one as well. Its definitions are still ambiguous, and the basic model rests entirely on expert opinion thus making the effectiveness ratings highly subjective. Finally, the system was originally designed to deal with driver trainers, and many of the questions (e.g., sound requirements) seem uniquely aimed at such devices, though with some modification, they might be made more universally applicable.

CARO'S METHOD

This method represents the first in a final set of models which do not depend upon the type of task analysis underlying the methods already described. Emphasis is on description of task elements and hardware rather than on an analysis of the psychological functions or behaviors demanded by the task. Description is at a more molecular level.

The first example of a method based on this kind of microanalytic approach is presented by Caro (1970), and is known as Equipment-Device
Task Commonality Analysis (TCA). This method has been designed to predict the potential transfer of training which may result when devices designed for one training situation are considered for use in another. The approach is based on Osgood's theory of transfer (see Section 3.1) in that the parameter of concern is the similarity in stimulus and response elements which exists between the training device and the operational situation.

The first step in the method consists of a description of the stimuli and the responses elicited by them in the operational equipment. This description focuses on both hardware (displays and controls) and non-hardware (environmental) stimuli. The second stage is a parallel analysis of the stimulus and response elements in the training device. In the third step the two lists of stimulus and response elements are compared, in order to identify for each element in the operational system an analogous element in the training device. For each match that is found, a rating of realism is obtained. For stimulus elements realism, relative to the real gear, is rated on a two-point scale in terms of both appearance and function. Each control or response element is rated for realism along five dimensions, including appearance, location, direction of movement, feel, and effect on displays.

In applying the method in any specific training context, the criterion performance to be trained is carefully analyzed to discover the critical stimulus components judged as necessary for satisfactory performance. These are flagged and are examined with respect to their judged realism in the training device being evaluated. As estimate of transfer for that device is obtained by considering the commonality (or similarity) between the two systems. If the stimuli and responses in the device are similar to those in the operational equipment, positive transfer is predicted; if the stimuli are similar and the responses are dissimilar, negative transfer is predicted.

In summary, this method represents a departure from the previous methods. It aims at prediction, even though it is at a very general level and would only allow for gross comparisons among devices. The prediction is aimed at transfer rather than acquisition and the inputs are task data at the molecular (display-control) level.
One problem with Caro's approach is that Osgood's transfer model is not entirely consistent with the available data. (See Bugelski & Cadwallader, 1956.) Similarly, we lack data which suggests how much commonality is sufficient for positive as opposed to zero transfer. Caro does not specify how the analyst solves this problem. Further, in analyzing stimuli, Caro has chosen the displays and controls as the level of analysis. Other levels of description might be used and, indeed, for Osgood, a display such as a radar screen would probably contain a large number of stimuli. In addition, this model only considers stimuli and responses, and does not allow for cognitive and other conceptions of transfer (such as transfer of principles, etc.).

QUANTITATIVE TASK DESCRIPTION

Rather different from the preceding systems is a method developed by Wheaton and Mirabella (Wheaton, Mirabella, & Farina, 1971; Wheaton and Mirabella, 1972; Mirabella and Wheaton, 1973). This scheme uses a detailed task description to generate data for use in a multiple regression model. The first step in this method consists of describing tasks at a detailed level by means of flow diagrams which indicate what the operator does and with which display and control elements. This information is used to calculate values on a set of quantitative dimensions which are, for the most part, hardware and engineering oriented. Specific values are obtained, for example, on such indices as the number of required primary responses, the number of contingency responses, the number of displays and controls used, and the Display Evaluation Index (DEI) (Siegel, Miehle, & Federman, 1962). These indices serve as the predictor variables.

Early work with the indices indicated that several were generic and could be applied to a broad range of tasks while others were useful only in more console-oriented or procedural tasks. It was also shown that the quantitative indices could be reliably applied and that they discriminated among sonar training devices.

In later studies with this method the investigators conducted acquisition and transfer research on a procedure-following task which could be physically manipulated to vary values of some of the indices. Such
manipulations did produce changes in acquisition performance which could be predicted by multiple regression equations based on the indices. Finally, it was demonstrated that the performance of subjects who transferred from one version of the task to another could also be predicted using either the predictor scale values on the acquisition task or the difference between predictor scale values on the training and transfer tasks.

In this system information derived from description of the tasks and equipment comprising a device are not used to determine training implications directly. Rather it provides a base on which the quantitative indices may be calculated, and then used to predict acquisition or transfer scores.

The approach is oriented toward prediction, is empirically based, and can be used to forecast both acquisition and transfer. Unfortunately, however, the method is still in its infancy. Many of the indices are applicable only in fairly proceduralized tasks, and more work is required on selecting and developing additional indices. The method would require a large amount of empirical work before really solid equations could be developed. Furthermore, the extremely detailed level of description which is used is viewed as a handicap in evaluating complex tasks or devices. Nevertheless, the general method appears to hold some promise for certain kinds of hardware-oriented tasks.

ALTMAN'S METHOD

The final model to be discussed is one whose development did not have the same impetus as the others. While not strictly concerned with training device effectiveness, Altman (1970) presented a microanalytic model that appears to be applicable for predicting transfer of training. His model includes considerations derived from Osgood's theory (i.e., stimulus and response similarity) as well as from Dallett's transfer model (emphasizing stimulus-response bond similarity). To apply Altman's formulations, the analyst first diagrams the operational and training tasks in terms of stimuli and responses. The following proportions are then calculated.

\[ P_r \] The probability that a response element in the transfer [operational] task will be in the original [training] task.
PS: The probability that a stimulus in the transfer task will be in the original learning task.

QR: 1 - PR

QS: 1 - PS

QRS: The probability that a response element is found in both tasks, bonded to [associated with or following] a stimulus element found in both tasks, but bonded with a different stimulus in the two tasks.

QSR: The probability that a stimulus element is found in both tasks, bonded with a response found in both tasks, but bonded with a different response in the two tasks.

Altman then postulates three kinds of transfer and defines them in terms of the following formulae:

Response Repertory Transfer (RR) = p^2

Response Association Transfer (RA) = p^2 (1 - 2[QRS + QS])

Stimulus Association Transfer (SA) = p^2 (1 - 2[QSR + QR]).

Net Transfer (NT), which would be of interest in device evaluation, would equal the sum of the three components:

NT = RR + RA + SA

The model is predictive, deals with transfer, and uses microanalytically derived inputs. At present, the most serious problem with this model appears to lie in defining and describing stimuli, responses, and bonds. The molecular level of description which is at least implied may not be practical when evaluating fairly complex training situations. The model has not been tested empirically, but this problem is not too critical, since it appears that it is, in fact, testable. However, as with Caro's and Wheaton and Mirabella's models, it ignores more complex mediators of transfer (e.g., knowledge), and does not consider the impact of acquisition effects (e.g., amount of original learning) on transfer. Finally, this approach, in its present form, does not consider the impact of other potentially important variables (i.e., such as training techniques) on acquisition or transfer.
CONCLUSIONS

The general conclusion to be drawn from the review of methods and models is that no existing model is entirely adequate for predicting the effectiveness of training devices. Several different approaches have been developed and all have suffered from a variety of serious weaknesses.

1. Most of the models were prescriptive rather than predictive, and were developed for specifying the design of training. Thus, they tended to focus on an analysis of the content of training but fell down in the specification of precise methods for implementing training.

2. Virtually all of the models had a scope limited to acquisition. None of the systems adequately considered both acquisition and transfer aspects of training device effectiveness.

3. There is a tendency to utilize a single level of description for input to the model (whether at the molar or molecular level). This limits the flexibility of the basic data to be processed and thus the output.

4. The definitions and procedures for data acquisition and processing tend to be complex, cumbersome and in many cases ambiguous.

5. A tendency exists to ignore the multidimensional nature of the problem and to oversimplify the approach by limiting the consideration to one or two dimensions (e.g., similarity, fidelity, etc.) thought to impact on transfer.

6. None of the methods is sufficiently concerned with the problem of quantification. Thus, none supplies acceptable and workable metrics for the crucial variables they consider, and this limits the form and usefulness of the outputs they provide.
7. Insufficient empirical support exists for both the underlying rationale and the procedures of the models.

However, taken as a group, they do identify the kinds of dimensions necessary for an adequate model, as well as the possible pitfalls in the development of an appropriate method. Examining the spectrum of systems and approaches from the literature, a multi-dimensional, multi-level model is suggested. Such a model should include:

1. Task analysis, at a gross behavioral level as well as a more molecular level. The gross analysis would be used to determine the behavioral or task communality in the training and operational situations, and to flag tasks which are critical for system performance (a la Chenzoff and Folley, and Shettel). The more molecular analysis would be used to examine the correspondence in representation of tasks present in both systems (cf. Caro). The task analysis would be orthogonal to other variables, i.e., each task would be considered separately under other dimensions of the analysis.

2. Acquisition analysis may be important as well. As emphasized in the analytic, non-predictive methods, one wants to train the trainee to do what he cannot do at the beginning of training. This would include an examination of the trainee's capabilities (skill or response repertory), and a comparison to the skills needed in the operational task. Additionally, one would want to single out difficult tasks, or tasks which require special training (e.g., special equipment, or part-task training, etc.) (cf. Chenzoff and Folley, and Shettel). Finally, the amount or stage of training necessary to achieve the desired transfer should be considered (cf. R. B. Miller, and Fitts).

3. Principles of learning and training techniques should be considered as they impact on acquisition of each kind
of task. Additionally, principles of transfer and techniques of training should be considered, in terms of the impact on transfer of various manipulations of acquisition and transfer conditions. These analyses would also be conducted task by task, so that possible interactions or lacks of generalizability may be detected.

4. Finally, the obtained information needs to be collated to predict training device effectiveness in terms of transfer of training. This synthesis is most easily accomplished when the level of measurement at each preceding stage is high. The synthesis would hopefully allow us to determine when particular dimensions of the model are or are not important, on the basis of their interactions with other aspects of the model.
THEORETICAL POSITIONS AND ISSUES

Having reviewed previous models and methods used to improve the effectiveness of training, attention can turn to a series of theoretical positions and issues. Consideration of these issues flows naturally from the review of previous models, inasmuch as each model has been based upon a number of assumptions (both explicit and implicit). Many of the models just reviewed assume, for instance, that transfer of training results from a kind of behavioral generalization. When confronted with two similar situations, the student will tend to behave in the second the same as he behaved in the first. A subsidiary assumption is that if he performed well in the first situation, he will perform well in the second.

The theoretical bases for these kinds of assumptions need to be examined prior to further modeling. This step is needed to insure a sound theoretical foundation for the preliminary predictive model, particularly in terms of choosing the constructs upon which to base the model. Toward this end the present section begins with a general review of two kinds of transfer theory. Theories based on similarity and mediation are examined with respect to their implications for a model. Description of these positions is then followed by a more detailed discussion of three pervasive issues in transfer of training which need to be considered in modeling. These include the importance of attempting to account for negative transfer effects, the role of fidelity in transfer, and the relation between transfer and amount and stage of practice.

SIMILARITY THEORIES OF TRANSFER

The degree of similarity between tasks is theoretically an important determinant of transfer between those tasks. In theory, the more closely tasks A and B are related, the more interaction there will be between learning A and learning B. The difficulty, however, lies in predicting whether the interaction will be favorable or unfavorable. This problem is essentially the same whether one is investigating the effects of task A on B or is interpolating task B learning between episodes of learning task A in order to find the effects of B on A.
In the retroaction (effects of B on A) literature it began to appear that overall similarity between, for instance, memorizing successive lists of quartermaster items for supply purposes, might have several different effects (Kling & Riggs, 1971). If the two lists were highly similar, there would be considerable facilitation. However, when the lists were slightly less similar (and the differences were appreciable to the experimenter), negative effects were obtained. Of course, if the differences were sufficiently great, the two tasks were neutral with respect to one another. The Skaggs-Robinson hypothesis (Robinson, 1927) was an attempt to display the apparent paradox as a theoretical graph, which is shown as Figure 6.

The problem had not arisen in the same form in the transfer (effects of A on B) literature, since an early development was the separation of similarity phenomena into input (S) and output (R) components. Poffenberger (1915) had shown that transfer between learning lists of noun and adjective pairs was positive when S and R remained the same, negative when R changed and S was the same, and zero when both S and R changed, from the first to the second task. Wylie (1919) demonstrated in addition, using maze-running techniques, that changes in S when R components remain the same also lead to positive transfer. Bruce (1933), using paired nonsense syllables, later confirmed these findings and showed that the relations still hold when similarity is substituted for sameness in S or R components. Many subsequent studies have shown the effects of gradients of similarity.

Several attempts have been made to incorporate the ways in which the amount of transfer varies with gradients of similarity into three-dimensional surfaces, relating S similarity on one axis and R similarity on another to amount and direction of transfer on the third. Best known of the similarity based transfer theories are those of Osgood, Dallett & Huston.

Osgood (1949) attempted to predict for both the retroaction and transfer situations by extrapolating the above and related findings in the three-dimensional surface shown as Figure 7. Stimulus relations between two tasks are depicted as varying from identical to neutral, while the responses continue through neutrality to opposition and antagonism. With identical stimuli, the effect of variation in required responses is depicted
Figure 6. The Skaggs-Robinson hypothesis.
Figure 7. Osgood's transfer and retroaction surface.

- I: identical
- S: similar
- N: neutral
- O: opposing
- A: antagonistic
as passing from maximum transfer at identity, through zero to negative transfer as antagonism is reached. With identical responses, transfer drops to zero as stimulus similarity decreases. With antagonistic responses, transfer rises to zero from negative as stimulus similarity decreases. The Osgood surface represents an attractive simplification and has become extremely well known.

The Osgood model, however, suffers both from errors of prediction and from the omission of other potentially relevant variables. Its major deficiency is the inaccuracy with which it predicts negative transfer, despite the fact that such prediction was the major reason for its construction. For example, it often occurs that rearranging the pairing of the old stimulus and response words may give rise to considerable interference (Porter & Duncan, 1953). Since the model has to be applied to individual word pairs, it cannot at the same time account for list structure, and therefore fails to predict the effect.

An extensive test of the Osgood surface was made by Bugelski and Cadwallader (1956). These investigators used four levels of S similarity and four levels of R similarity, testing all sixteen resulting relationships. Both transfer and retroaction measures gave similar results. By and large, the results were in accordance with theory for positive transfer, but did not show the predicted gradient of negative transfer. In fact, the resulting curve of transfer against R similarity contains a reversal. The data are shown superimposed (dotted lines) on the Osgood surface in Figure 7. By comparison with Figure 6, it can be seen that what has reappeared is the Skaggs-Robinson paradox. Wimer (1964) has obtained similar findings. The negative transfer problem has not therefore been resolved; the issue is considered further in a later portion of this section.

In order to deal with re-paired lists, where the rearranged items might be similar rather than identical with the original items, Dailet (1965) has constructed another surface, which complements Osgood's; it considers stimulus and response similarity effects on transfer when responses are rearranged (re-paired with different stimuli) in task B. Since the surface is complementary, it cannot deal with the relationships considered by Osgood, and is equally deficient with respect to the additional
transfer variables described below. Dallett's surface has received little experimental attention.

Both Osgood's and Dallett's models omit the similarity relations between the stimulus words in task A and the response words in task B, and between the A responses and B stimuli. Houston (1964) has constructed a further surface, which uses these relations instead of Osgood's response similarity and stimulus similarity axes. Unfortunately, his model has also received little experimental attention.

MEDIATION THEORIES OF TRANSFER

The transfer surfaces make the assumption that decreasing the input or output similarity will yield transfer relationships differing only by degree from the relationships predicted on the basis of stimulus or response identity. It does appear to be broadly true that positive transfer relations vary in proportion to stimulus generalization gradients (Gibson, 1941). Thus, trainees who have learned to report "target detected" on discriminating a 500 Hz tone are also likely to respond to a 250 Hz or 1,000 Hz tone, and are therefore likely to transfer well to a task incorporating these stimulus frequencies. However, generalization need not be based directly on a psychophysical dimension. Obviously, men will react to the command "stop" with almost the same alacrity as in responding to "halt"; in this case the similarity is a learned semantic relation rather than a direct physical similarity. Mediation is therefore an intervening mechanism which makes transfer possible. It is treated below in terms of the "transfer of principles" and "hypothesis theory."

The most abstract form of mediated transfer is provided by the transfer of principles. Learning that, for instance, "light is refracted towards the perpendicular as it moves from a less dense to a denser medium" is potentially useful in transfer among an infinite variety of aiming tasks. There is, however, the difficulty that the formulation may be too abstract for the trainee to apply it usefully, and the concomitant problem that the trainee has to spend time in learning a verbal formulation rather than in actual practice of an aiming task.

Although Judd's (1908) famous experiment showed that boys were aided in throwing darts at submerged targets by instruction in the principles of
light refraction, Coleville (1957) showed that an equivalent amount of
time spent in practicing games skill produced the same effects. The problem
of formulations too abstract to apply was illustrated by Waters (1928), who
compared different methods of training for successful transfer between
different forms of the game of "nim." The general formulation "always
draw so as to leave a multiple of three" gave better transfer, despite the
fact that it was untrue for the more complex problems, than the truly
general formulation "always draw so as to leave a multiple of the sum of
the highest and lowest draws." Most successful of all was the method of
drawing the trainee's attention to the relevant cues, merely by requiring
him to call out how many beads were left at the end of each draw.

While the work in this area is fragmentary and incomplete, Holding's
(1965) survey appears to lead to a conclusion based on "need to know."
The trainee should not be loaded with a burden of verbal instructions or
abstract principles if these are unnecessary. He should be given only
what he needs to know at his current stage of training, in the light of what
he will be required to do in the operational task. When instructions or
principles are communicated, they should be as concrete, simple, and direct
as possible.

At certain junctures in a learned task, or in the acquisition of novel
tasks, or on performing certain kinds of transfer tasks, the trainee will be
confronted with a choice of actions. In these circumstances his behavior
will be based on the testing of hypotheses. The relevant hypothesis theory
arises in the literature of concept formation and problem-solving, where Levine
(1970) introduces the notion of subset sampling.

The suggestion is that a group of related hypotheses form a class, or
"domain", whose members are sampled together or in close succession with
results which affect the probability of the class as a whole. This leads
to a reduced likelihood of solution for problems requiring the testing of
hypotheses which lie outside the sampled subset. In the task explored by
Fingerman (1972), a finite subset was formed by presenting stimulus cards
in which shape, color, added bar, and size of symbol were varied. Correct
responding was made contingent on one of shape, color or bar over a series
of concept problems; symbol size was an irrelevant variable until the critical,
final problem was introduced. As the series of problems proceeded, blank trials showed that the size hypothesis was at first sampled increasingly often, since the other hypotheses in the same subset were being reinforced. With more problems the size hypothesis, which was never itself reinforced dropped in probability to approximately zero, with the result that the final size-solution problem was less readily solved than even the first of the other problems.

The way in which transfer is affected by hypothesis formation has been defined by Levine (1973) as follows: "When S receives a series of problems, he infers from the first n solutions the domain within the universe from which the n + 1st solution will be taken. He will start the n + 1st problem by sampling Hs from this domain." If sampling a related hypothesis is appropriate, there will be high positive transfer; if, on the other hand, conditions are so arranged that a related hypothesis is inappropriate, there will be apparent negative transfer. It is clear, therefore, that in the absence of clear indication that what is required is different, trainees will continue to behave in what may become inappropriate ways, as a result of prior training.

THEORETICAL ISSUES IN MODELING

The major theories just described raise a number of more specific theoretical issues. These issues have implications for the manner in which transfer of training in the military setting is viewed, and for the development of a model which can be used to predict such transfer. Three considerations will be discussed, including: 1) the nature of negative transfer and its permanence; 2) concepts of similarity; and 3) the role of amount and stage of training in transfer.

Negative Transfer. Negative transfer does not necessarily imply a reversal of behavior. Rather, it may be viewed as positive transfer in an inappropriate context. Trainees will tend to make an old response in a new situation, to the extent that they fail to distinguish between the old and new situations, and between the old and new responses. Failures to distinguish are thus more probable when the stimulus aspects of tasks A and B are similar, and when the required responses are similar (but are scored differently). It therefore appears that the Skaggs-Robinson hypothesis is
essentially correct, although oversimplified, while the Osgood theory is misconceived.

With respect to the stimulus situation, it is clear that not all the stimuli in tasks A and B can be identical. For practical purposes, it will commonly occur that all or most of the task stimuli are the same, but that either the contextual cues are different or the old and new situations are differentiated by an initial cue or instruction. The amount to which situations A and B are confused will then depend upon the number and prominence of these extra-task cues. An initial cue or instruction imposes a memory load and is consequently fallible, leading to frequent lapses. Contextual cues in the two tasks may vary from being minimally different, allowing appreciable interference in transfer, to offering cue differences which constantly interact with the ongoing task activity. For instance, there is surprisingly little negative transfer from driving on the right to driving on the left of the road, if the driving is done in an indigenous automobile. The transposed position of the driving seat and steering wheel (the contextual cues) act as constant reminders of the "scorable" differences between the two tasks.

On the response side, there is ample evidence that the requirement of different responses between two tasks does not of itself make for negative transfer, either in verbal tasks (Kling & Riggs, 1971) or in motor skills (Bilodeau & Bilodeau, 1961). When negative transfer is found at all, it tends to occur when the responses in the new and old situations differ, but the difference is not perceivable by the learner. An example is afforded by the negative transfer from upward to downward lever movement studied by Adams (1954). In contrast, one would not expect negative transfer from task A, learning to run ten paces on the word "go", to task B, learning to pull a parachute ripcord on the word "go." What matters is the degree of response generalization, suggesting that Osgood's "antagonistic" responses, like elated versus dejected, should be placed at the "similar" rather than the "different" end of the response continuum. The Bugelski-Cadwallader data plotted in Figure 7 show a clear decrease in negative transfer as responses change from similar to opposed (following the initial drop due to the change of scoring criterion).
The usual view that negative transfer increases, when the stimuli are the same, as the responses in the two tasks differ, should therefore be reversed. It might be suggested instead that negative transfer increases, when the stimuli are barely different, as the responses become more similar. The statement then bears some resemblance to the statement for positive transfer, except that the scoring criterion is reversed. A plausible alternative to the Osgood surface is presented as Figure 8, by way of illustration. It would probably apply only to initial transfer and, in practical training, only to one or two of the component responses in the training activity.

Given, then, that negative transfer may occur under some conditions, its duration is of concern, particularly in the applied realm of military training. Total negative transfer, in the sense of long-term retardation of the learning process in task B, is rarely found. It can of course be contrived by structuring the situation in such a way that the difference cues between tasks A and B are minimally obvious to the subject. More often, since the requirement for negative transfer is that tasks A and B should be highly similar, the relationship will also offer a potential for positive transfer. In any event, it seems probable that, during the course of task B learning, negative transfer may rapidly give way to positive transfer as the trainee acquires discriminative control of the two responses.

Occasional lapses, or intrusive errors, are often found and will be important in certain kinds of tasks. Clearly, a few errors carried over by habit interference from training routines to the early trials of an operational task will not matter unless the equipment is delicate or dangerous, or unless the consequences of an error are of greater importance than the eventual acquisition of skill. Carrying over a tendency to give the radio call-sign used in training instead of the aircraft call-sign will make for easily correctible errors, and can be tolerated if the overall transfer from the training system is positive. However, in the training of bomb disposal crews, it is advisable to avert intrusive errors on the first, and every subsequent, trial of the transfer task. The evaluation of a training system for its negative transfer potential must therefore be far more stringent in these circumstances. Thus, it is important for a transfer model to consider
Figure 8. An alternative to the Osgood surface.

I identical
S stimulus
S similar
N neutral
R response
O opposing
A antagonistic
the type of transfer appropriate for the operational situation. In some instances first-trial transfer may be of concern; in other cases a measure of overall transfer may be more meaningful.

**Fidelity.** Just as there are alternative ways of viewing transfer, so are there several different ways of conceiving of similarity. It has been seen that, theoretically, the transfer relation between two tasks is crucially dependent in various ways upon the degree of similarity between them. However, task similarity may not coincide with equipment fidelity. The term "fidelity" was originally intended to refer to similarity in the engineering context, as distinct from psychological similarity. Thus, when Grimsley (1969) concludes that low fidelity is no handicap to the effectiveness of training devices, equipment fidelity is meant; it is not suggested that high similarity, between what the trainee does in the training and in the transfer tasks, is unnecessary. In contrast, when Briggs & Johnston (1966) conclude that high S fidelity is necessary in procedural training, while high R fidelity is necessary for motor skill acquisition, it is clear that neither statement refers directly to engineering hardware.

Deviations from equipment fidelity will hypothetically affect transfer to the extent that inadequate, or different, displays lead trainees to attend to the wrong perceptual cues, or that inappropriate controls lead to practicing wrong movement sequences. As noted above, interference arising from those sources may be transitory, or limited to isolated, intrusive errors whose seriousness will depend upon the nature of the operational task. Intrusive errors are more likely to be carried over from first-task learning when the initial training is extended, and less when it is moderate. The amount of first-task learning is also, therefore, an issue of theoretical importance.

**Amount and Stage of Training.** Thus far it has been assumed that the level reached in training on task A has been standard, while various intertask relationships between task A and task B have been discussed. However, it cannot be simply assumed that increasing or decreasing the amount of task A training has a direct proportional effect upon the other transfer variables. The amount-of-first-task-training variable has been inadequately
researched to date, but the following sections attempt to deal with the major issues.

Most theories agree that as training approaches asymptote, the more skills and abilities are available for transfer. If the input and output characteristics of the training devices do not misrepresent the characteristics of the operational task, such transfer will be positive; if not, negative tendencies will arise. However, with overlearning beyond asymptote, the danger is that trainees may begin to learn detailed idiosyncrasies of the training device which will cause interference on transfer to the real equipment (Weitz and Adler, 1971).

Where transfer is made from an incomplete stage of training, other difficulties arise. It has been hypothesized that the relevant similarity variable is functional, or psychological, similarity, and that this depends upon mediate or immediate stimulus and response generalization. It appears, however, that the development of generalization and differentiation does not follow a simple time course. Gagné & Foster (1949) have explored this effect in discrete motor tasks, finding that the tendency to generalize from the task stimuli to other stimuli at first increases over training trials, later diminishing as discrimination develops. It will be recalled that a parallel effect was pointed out in hypothesis theory for concept tasks (Fingerman, 1972). If this is generally true, it suggests that the greatest interference will be carried over from an intermediate level of training. This appears to be true, both for verbal learning (Bugelski, 1942) and for motor learning (Mandler, 1954).

Equivalent effects may occur in learning task B. Thus, in Adams' (1954) study, negative transfer did not arise until after six or seven trials had been conducted on the second task. Taking both effects together, it would be expected that maximum interference would take place when both tasks A and B are at early stages of practice. Such an effect was in fact reported early in the literature of transfer (Snipola & Israel, 1933), and has never been experimentally refuted. It appears reasonable to assume that maximum interference will arise at the stage when the input and output variables in both tasks are at a low level of discriminability.
The preceding discussion is directed at the substantive phase of skill learning, where input and output relations are becoming correctly appreciated and implemented; the task is, as it were, being "put together" by the trainee. Fitts & Posner (1967) refer to this as the "associative" stage, and suggest that it is preceded by a "cognitive" stage, during which the trainee is finding what the task is about, and followed by an "autonomous" stage in which performance becomes automatic.

During the "cognitive" stage, the trainee is preoccupied with input variables. What he is required to process ranges from instructions and demonstrations to perceptual inputs from the equipment layout and workspace arrangement and possibly also the terrain and general environmental cues. It appears probable, therefore, that stimulus fidelity requirements will be paramount in the very early stages of training. On the other hand, in the later stage, the trainee is becoming automatic in his control movements and less responsive to external inputs. In these circumstances it must follow that response fidelity requirements should be stringent.

SUMMARY

In the survey of previous methods and models, two distinct kinds of approaches were uncovered. In one, emphasis was placed upon training for acquisition of skill, the assumption being that the skill would carry over to the (similar) operational situation. In the other, the focus was on establishing the degree of similarity between the training and operational contexts as a basis for predicting transfer. Not surprisingly, these two approaches were broadly mirrored in the mediation and similarity theories reviewed in the present section.

Both theoretical positions are of interest. The more viable theory for modeling purposes would appear to be the similarity theory. This is true for the primary reason that it may simply be easier to operationally define similarity. However, it must be recognized that current similarity models seem incapable of accommodating all of the gradients of similarity which are possible. What is worse, these similarity relations by no means exhaust the list of relevant and powerful variables. Similarity may be a necessary condition for transfer to occur, but it surely is not sufficient.
In addition, the similarity theories and surfaces have received their impetus from the verbal learning paradigm. Generalization of results from this paradigm to tasks of interest in the applied military training context is often difficult as discussed in the next section.

The mediational approach is important as well, since it further defines what conditions are necessary for transfer. The emphasis is on particular skills or knowledge "carrying over" to a transfer task; the skills and knowledge are thus mediators. This approach recommends that, for modeling purposes, some attention be paid to such skills and knowledge. An analysis to discover necessary mediational elements would help to define acquisition requirements for training. Once again, however, acquisition related to mediators may be a necessary but not a sufficient element for the occurrence of transfer.

Neither of these approaches includes all the variables known to impact on transfer. For example, all kinds of tasks, from learning the names of regiments and corps to radar servicing and helicopter flying, show the dependence of transfer upon the amount and stage of prior training. For verbal tasks, Kling & Riggs (1971) conclude that there is experimental support for requiring transfer predictions to take into account: response learning; response differentiation; stimulus differentiation; forward associations; backward associations; and list differentiation. For perceptual-motor tasks, it has been shown that a cluster of relations between the transfer and transferred tasks other than similarity, grouped under the heading of "task difficulty" may have important consequences for transfer.

It should be borne in mind that the scope of the present section was limited to consideration of major variables of theoretical interest, although a very large number of substantive factors will enter in varying degrees to military applications of transfer theory. These relationships are discussed in the next section of the report which deals with substantive issues in transfer and their empirical support.
SUBSTANTIVE ISSUES

OVERVIEW

The purpose of this section of the report is to describe and synthesize the data in a set of empirical studies dealing with the effects of specific variables on transfer of training. This information, coupled with that obtained from previous approaches as well as theory, will contribute to the structure and content of a model for predicting training device effectiveness.

From the large amount of empirical research surveyed in this and other literature reviews, many variables influencing transfer of training have been identified. Given this wealth of information, however, it is somewhat surprising that relatively little systematic knowledge exists concerning ways to promote efficient transfer through appropriate design of the training device or overall training system.

In retrospect, there are at least three major reasons for this situation. First, much of the empirical research on transfer has been done within the context of the verbal learning paradigm. In this kind of research, the transfer and training tasks, and the associated stimuli and responses are precisely specified. Furthermore, variables which are difficult to measure in more practical and complex situations are rather easier to scale, and the influences of non-manipulated variables are likewise easier to control. Finally, it is difficult at best to generalize findings based on the verbal learning paradigm to tasks of more direct relevance to military training.

A second reason for the general lack of a systematic body of knowledge is that most of the research has been directed toward the testing of theoretical hypotheses. In this research the focus of interest is typically not on practical implications, but rather on the confirmation of an experimental effect. As a corollary, this strategy of verifying micro-theories of transfer usually involves restrictive stimulus materials, limited response dimensions, and so forth. While this type of research is critically important and necessary, it greatly increases the difficulty of organizing and synthesizing the data into a useful structure or macro-theory of transfer of training.
Finally, a third major problem concerns the difficulty in generalizing a given experimental finding from a situation in which selected variables have been held constant to situations where those same variables may vary drastically. For example, Cox, Wood, Boren, & Thorne (1965) studied the effect of fidelity of simulation on the transfer of training of a 92-step procedural task. Twelve different training devices were used representing variations in device "functional and appearance" fidelity. Fidelity was manipulated by using different mock-ups of a control panel varying in: 1) size and type of housing; and 2) panel representation (i.e., hot panel, cold panel, photograph of the panel, etc.). The major finding was that:

"When men are being trained to perform a fixed procedure, the requirements for functional fidelity in the training device are quite low. A line drawing of the man-machine interface will train men as effectively in this circumstance as will a device of higher fidelity" (p. vi).

Given this seemingly straightforward finding, the question arises as to how far to generalize this result. Type of task seems to be an important qualification inasmuch as the authors are willing to state their conclusion for a "fixed procedure" task, but not for other types of procedural tasks nor, for that matter, for any other type of task. Since the subject population was fixed (i.e., trainees) the question arises as to whether or not the result would hold for other types of subjects (e.g., experienced operators). Similarly, the training program was held constant and was delivered by proficient instructors. Would the same result, therefore, hold for different training techniques, or with different instructors? Finally, would the result still be valid were other loci on the fidelity dimensions examined? All of these questions are concerned only with main effects, the number of possible interactions influencing transfer being enormous. Likewise, other potentially potent variables, known and unknown, were not considered. In attempting to apply the findings of such a study, therefore, one clearly must be aware of these kinds of confounding and the limitations which they impose on generalization of results.
In spite of the kinds of problems just alluded to, an attempt has been made in the present report to document, organize, and synthesize information about the impact of specific variables on transfer of training. For some variables, extensive data was available and synthesis proved to be feasible. For many other variables, however, valid conclusions could not, in good conscience, be drawn from the few or conflicting studies available.

In subsequent portions of this section the influence of numerous variables on transfer is documented and summarized. The data are presented for each of three major classes of variables, including: 1) those influencing the efficiency of training (training variables); 2) those influencing the appropriateness of training (device variables); and 3) those dealing with aspects of the content of training (task variables). In the final portion of this section an attempt is made to synthesize data cutting across all three categories. This is accomplished by presenting a series of training principles designed to promote effective transfer.

**ORGANIZATION OF DATA**

One way of organizing information about the impact of substantive variables on transfer is to determine the proportion of studies which have dealt with the same issue. The assumption is that, if several researchers have considered a variable important enough to investigate experimentally, that variable is correspondingly important for transfer. Working on this premise, a subset of 89 studies was selected, reviewed, and cross-tabulated. The specific set chosen uniformly met a number of stringent criteria related to rigor of experimental design. Each study dealt empirically with a relatively specific variable or set of variables.

Tables 4 and 5 cross-index the type of variable(s) investigated in each study, with a list of the related studies. For the convenience of the reader, the variables and the references are numbered. The numbers following each of the variables refer to the relevant reference; the numbers following the references refer to the relevant variables. Most studies included in this review deal with more than one variable. This is indicated when a reference has more than one variable number following it. Thus, when counting the number of studies dealing with a particular variable, the total exceeds the number of references. Complete citations for these studies
Table 4

SUBSTANTIVE VARIABLES INDEXED BY RELEVANT DOCUMENTS

<table>
<thead>
<tr>
<th>Variable</th>
<th>Relevant Documents (from Table 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Amount of Practice (First Task Mastery)</td>
<td>6, 11, 17, 28, 38, 50, 59, 67, 70, 74, 76, 87, 89</td>
</tr>
<tr>
<td>2. Augmented Feedback</td>
<td>11, 22, 28, 29, 31, 33, 46, 48, 49, 59, 60, 72, 79, 85, 86</td>
</tr>
<tr>
<td>3. Control Parameters</td>
<td>8, 23, 28, 29, 32, 37, 38, 43, 55, 72, 74</td>
</tr>
<tr>
<td>4. Device Characteristics and Utilization</td>
<td>11, 12, 15, 20, 24, 34, 35, 44, 47, 62, 69, 75, 84, 85</td>
</tr>
<tr>
<td>5 Display-Control Relationship (Compatibility)</td>
<td>1, 9, 16, 23, 26, 28, 37, 38, 43, 74</td>
</tr>
<tr>
<td>6. Fidelity (Similarity):</td>
<td></td>
</tr>
<tr>
<td>a. Environmental</td>
<td>5, 13, 14, 29, 30, 31, 60, 65, 69, 75, 76, 84, 85</td>
</tr>
<tr>
<td>b. Response</td>
<td>1, 3, 9, 14, 16, 23, 25, 26, 28, 29, 34, 38, 41, 42, 49, 53, 55, 67, 68, 74, 84, 87</td>
</tr>
<tr>
<td>c. Stimulus</td>
<td>1, 3, 5, 13, 14, 16, 23, 25, 26, 28, 29, 30, 31, 34, 38, 41, 42, 48, 49, 53, 55, 57, 60, 63, 64, 66, 67, 68, 69, 74, 75, 84, 85, 87, 88</td>
</tr>
<tr>
<td>7. Knowledge of Results</td>
<td>27, 46, 60, 72, 86, 88</td>
</tr>
<tr>
<td>8. Motion Simulation</td>
<td>13, 29, 69</td>
</tr>
<tr>
<td>9. Previous Experience (Trainees)</td>
<td>28, 37, 51, 59, 60</td>
</tr>
<tr>
<td>10. Stimulus Predifferentiation</td>
<td>4, 17, 22, 25, 39, 51, 63, 64, 66, 73, 84</td>
</tr>
<tr>
<td>11. Stimulus-Response Associations</td>
<td>1, 3, 16, 23, 25, 26, 38, 50, 63, 77</td>
</tr>
<tr>
<td>12. Stimulus Variability</td>
<td>5, 30, 48, 57, 60, 67</td>
</tr>
</tbody>
</table>
### Table 4 (Cont'd)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Relevant Documents (From Table 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>13. Student Characteristics</td>
<td>5, 6, 21, 28, 40, 51, 61, 71, 81, 82, 83</td>
</tr>
<tr>
<td>15. Task Duration</td>
<td>5</td>
</tr>
<tr>
<td>16. Task Organization</td>
<td>19, 63, 66, 73, 77</td>
</tr>
<tr>
<td>17. Tasks:</td>
<td></td>
</tr>
<tr>
<td>a. Task Analysis</td>
<td>8, 10, 17, 21, 47, 53, 58, 61, 64, 78, 83</td>
</tr>
<tr>
<td>b. Tasks, Motor</td>
<td>1, 5, 9, 23, 25, 26, 32, 38, 41, 42, 43, 49, 54, 55, 67, 68, 72, 74, 76, 86, 87</td>
</tr>
<tr>
<td>c. Tasks, Schema Learning</td>
<td>6, 7, 17, 37, 39, 45, 50, 51, 59, 63</td>
</tr>
<tr>
<td>d. Tasks, Sequential</td>
<td>2, 14, 19, 34, 66, 73</td>
</tr>
<tr>
<td>e. Tasks, Time-Sharing</td>
<td>2, 9, 84</td>
</tr>
<tr>
<td>18. Training, Adaptive</td>
<td>11, 22, 85</td>
</tr>
<tr>
<td>19. Training Requirements</td>
<td>17, 27, 65, 73, 86</td>
</tr>
<tr>
<td>20. Whole-Part (Part-Whole)</td>
<td>2, 17, 19, 73, 77, 84</td>
</tr>
<tr>
<td>Transfer</td>
<td></td>
</tr>
</tbody>
</table>
## Table 5
RELEVANT DOCUMENTS INDEXED BY SUBSTANTIVE VARIABLES

<table>
<thead>
<tr>
<th>Document</th>
<th>Variables (from Table 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Adams, J.A. (1954)</td>
<td>5, 6b, 6c, 11, 17b</td>
</tr>
<tr>
<td>3. Altman, J.W. (1970)</td>
<td>6b, 6c, 11</td>
</tr>
<tr>
<td>4. Arnoult, M.D. (1957)</td>
<td>10</td>
</tr>
<tr>
<td>5. Baldwin, R.D., &amp; Wright, A.D. (1961)</td>
<td>6a, 6c, 12, 13, 15, 17b</td>
</tr>
<tr>
<td>8. Brecke, F., &amp; Gerlach, V. (1972)</td>
<td>3, 17a</td>
</tr>
<tr>
<td>13. Cohen, E. (1970)</td>
<td>4, 6a, 6c, 8</td>
</tr>
<tr>
<td>14. Cox, Wood et al. (1965)</td>
<td>4, 6a, 6b, 6c, 17d</td>
</tr>
<tr>
<td>15. Crook, W.G. (1967)</td>
<td>4</td>
</tr>
<tr>
<td>16. Day, R.H. (1956)</td>
<td>5, 6b, 6c, 11, 14</td>
</tr>
<tr>
<td>17. Deno, S.L. et al. (1971)</td>
<td>1, 10, 17a, 17c, 19, 20</td>
</tr>
<tr>
<td>18. Ellis, H.C. (in Marx, 1969)</td>
<td>A summary of several variables</td>
</tr>
<tr>
<td>19. Elmos, D.G. et al. (1972)</td>
<td>16, 17d, 20</td>
</tr>
<tr>
<td>20. Flexman, R.E. et al. (1953-54)</td>
<td>4</td>
</tr>
<tr>
<td>21. Fox, W.L. et al. (1969)</td>
<td>13, 17a</td>
</tr>
</tbody>
</table>

55
Table 5 (Cont'd)

<table>
<thead>
<tr>
<th>Document</th>
<th>Variables (from Table 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>22. Franken, R.E. et al. (1971)</td>
<td>2, 10, 18</td>
</tr>
<tr>
<td>23. Gagné, R.M. (1948)</td>
<td>3, 5, 6b, 6c, 11, 17b</td>
</tr>
<tr>
<td>25. Gagné, R.M., Baker, K.E., &amp; Foster, H. (1950: &quot;Transfer of ...&quot;)</td>
<td>6b, 6c, 10, 11, 17b</td>
</tr>
<tr>
<td>28. Geiselhart, R. (1966)</td>
<td>1, 2, 3, 4, 5, 6b, 6c, 9, 13</td>
</tr>
<tr>
<td>29. Gerathewohl, S.J. (1969)</td>
<td>2, 3, 4, 6a, 6b, 6c, 8</td>
</tr>
<tr>
<td>30. Hammerton, M. (1966)</td>
<td>6a, 6c, 12</td>
</tr>
<tr>
<td>31. Hammerton, M., &amp; Tickner (1967)</td>
<td>2, 6a, 6c</td>
</tr>
<tr>
<td>34. Isley, R.N. (1968)</td>
<td>4, 6b, 6c, 17d</td>
</tr>
<tr>
<td>37. Kelley, C.R. et al. (1966)</td>
<td>3, 5, 9, 17c</td>
</tr>
<tr>
<td>38. Kendrick, H.W. (1971)</td>
<td>1, 3, 5, 6b, 6c, 11, 17b</td>
</tr>
<tr>
<td>41. Leonard, S.D., Karnes, E.W. et al. (1970)</td>
<td>6b, 6c, 14, 17c</td>
</tr>
<tr>
<td>42. Lewandowski, J.T. (1971)</td>
<td>6b, 6c, 14, 17b</td>
</tr>
</tbody>
</table>

56
Table 5 (Cont'd)

<table>
<thead>
<tr>
<th>Document</th>
<th>Variables (from Table 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>43. Lewis, D., &amp; Shepard, A.H. (19...</td>
<td>3, 5, 17b</td>
</tr>
<tr>
<td>44. Little, A.D. (1973)</td>
<td>4</td>
</tr>
<tr>
<td>45. Logan, T.H., &amp; Wodtke, K.H. (1968)</td>
<td>17c</td>
</tr>
<tr>
<td>46. Lucas et al. (1973)</td>
<td>2, 7</td>
</tr>
<tr>
<td>47. Lumsdaine, A.A. (1960)</td>
<td>4, 17a</td>
</tr>
<tr>
<td>48. Mackie, R.R., &amp; Harabedian, A. (1964)</td>
<td>2, 6c, 12</td>
</tr>
<tr>
<td>49. MacRae, A.W., &amp; Holding, D.H. (1966)</td>
<td>2, 6b, 6c, 14, 17b</td>
</tr>
<tr>
<td>50. Mandler, G. (1962)</td>
<td>1, 11, 17c</td>
</tr>
<tr>
<td>52. McQuarrie, D., &amp; Grotelueschen, A. (1971)</td>
<td>17c</td>
</tr>
<tr>
<td>53. Micheli, G.S. (1972; TAEG 2)</td>
<td>6b, 6c</td>
</tr>
<tr>
<td>54. Moncrieff, J. (1966)</td>
<td>14, 17b</td>
</tr>
<tr>
<td>55. Muckler, F.A., &amp; Matheny, W.G. (1954)</td>
<td>3, 6b, 6c, 14, 17b</td>
</tr>
<tr>
<td>56. Murdock, B.B. (1957)</td>
<td>A summary of several variables</td>
</tr>
<tr>
<td>57. Murphy, W.W. (1971)</td>
<td>6c, 12, 14</td>
</tr>
<tr>
<td>58. Nehnevajsa, J. et al. (1960)</td>
<td>17a</td>
</tr>
<tr>
<td>59. Neiberg, A.D. (1968)</td>
<td>1, 2, 9, 17c</td>
</tr>
<tr>
<td>60. Overing, R.L., &amp; Travers, R.M. (1967)</td>
<td>2, 6a, 6c, 7, 9, 12</td>
</tr>
<tr>
<td>62. Poe, A.C., &amp; Lyon, V.W. (1952)</td>
<td>4</td>
</tr>
<tr>
<td>63. Price, R.H., &amp; Hill, R. (1968)</td>
<td>6c, 10, 11, 16, 17c</td>
</tr>
<tr>
<td>64. Price, H.E., Older, H.J. et al. (1957)</td>
<td>6c, 10, 17a</td>
</tr>
<tr>
<td>65. Prophet, W.W. (1963)</td>
<td>6a, 19</td>
</tr>
<tr>
<td>Document</td>
<td>Variables (from Table 4)</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>66. Reynolds, J.H. (1968)</td>
<td>6c, 10, 16, 17d</td>
</tr>
<tr>
<td>67. Rivenes, R. (1967)</td>
<td>1, 6b, 6c, 12, 14, 17</td>
</tr>
<tr>
<td>68. Rivenes, R., &amp; Caplan, C.S. (1972)</td>
<td>6b, 6c, 14, 17</td>
</tr>
<tr>
<td>69. Rolfe, J.M. et al. (1970)</td>
<td>4, 6a, 6c, 8</td>
</tr>
<tr>
<td>70. Roscoe, S.N. (1971)</td>
<td>1</td>
</tr>
<tr>
<td>72. Salvendy, G., &amp; Harris, D.R. (1973)</td>
<td>2, 3, 7, 17</td>
</tr>
<tr>
<td>73. Schwartz, R.M., &amp; Humphrey, M.S. (1973)</td>
<td>10, 16, 17d, 20</td>
</tr>
<tr>
<td>74. Shephard, A.H., &amp; Lewis, D. (1950)</td>
<td>1, 3, 5, 6b, 6c, 17b</td>
</tr>
<tr>
<td>75. Shettel, H.H. et al. (1971)</td>
<td>4, 6a, 6c, 19</td>
</tr>
<tr>
<td>76. Singer, R. N. (1966)</td>
<td>1, 6a, 14, 17b</td>
</tr>
<tr>
<td>77. Slamecka, N.J. et al. (1972)</td>
<td>11, 16, 20</td>
</tr>
<tr>
<td>78. Smith, B.J. (1965)</td>
<td>17a</td>
</tr>
<tr>
<td>79. Smode, A.F. (1972)</td>
<td>2</td>
</tr>
<tr>
<td>80. TAEG 1 (1972)</td>
<td>A summary of several variables</td>
</tr>
<tr>
<td>81. Tallmadge, G.K. (1968)</td>
<td>13</td>
</tr>
<tr>
<td>82. Tallmadge, G.K., &amp; Shearer, J.W. (1968)</td>
<td>13</td>
</tr>
<tr>
<td>83. Tallmadge, G.K. et al. (1957)</td>
<td>13, 17a</td>
</tr>
<tr>
<td>84. Valverde, H.H. (1968)</td>
<td>4, 6a, 6b, 6c, 10, 17e, 20</td>
</tr>
<tr>
<td>85. Voss, H.A. (1969)</td>
<td>2, 4, 6a, 6c, 18</td>
</tr>
<tr>
<td>87. Weitz, J., &amp; Adler, S. (1971)</td>
<td>1, 6b, 6c, 14, 17b</td>
</tr>
<tr>
<td>88. Wiener, E.L. (1967)</td>
<td>6c, 7</td>
</tr>
<tr>
<td>69. Williges, R.C., &amp; Baron, M.L. (1973)</td>
<td>1, 14</td>
</tr>
</tbody>
</table>
The choice of labels for the variables listed in Table 4 has been made somewhat arbitrarily. Unfortunately, there are few common usages among authors. For example, "fidelity" has often been used as a direct synonym for "similarity," despite the distinction between these two terms described earlier. In subsequent discussions, an attempt is made to amplify the meanings of these variable labels. In any event, the twenty major variables which have been identified fall rather conveniently into three general classes: 1) "Training" variables, 2) "Device" variables, and 3) "Task" variables. Variables in the first category (training variables) can be viewed as affecting the "efficiency" with which any task could be trained. For the most part, the variables are basically independent of a specific device implementation. This category includes:

1. Amount of practice,
2. Knowledge of results,
3. Previous specific experience of the trainees,
4. Trainee characteristics,
5. Training requirements,
6. Part-whole training,
7. Augmented feedback,
8. Adaptive training, and

The second category, device variables, consists of studies which have looked at the "appropriateness" of the representation of the task to be trained. Typically, this issue has been approached along some fidelity or similarity dimension. Thus, the variables in this category include:

1. Fidelity (environmental, stimulus, response),
2. Control parameters,
3. Device characteristics and utilization patterns,
4. Motion simulation,
5. Display-control relationships, and
The third category consists of studies dealing with specific task variables. These variables are:
1. Task difficulty,
2. Task duration,
3. Task organization,
4. Stimulus variability, and
5. Task analysis.

Originally included in this category was "type of task"; however, this proved to be primarily a between-study variable. In those cases where type of task turned out to be a within-study variable, the study was listed under the task analysis heading.

CATEGORY I - TRAINING VARIABLES

Amount of Practice. This variable has been called "degree of original learning" and "first-task mastery" in addition to "amount of practice." In Table 4, 13 studies were listed as relevant to this topic. The common manipulation was some specified performance criterion of first-task learning or practice, typically either number of trials or a pre-set performance level. For example, Shepard and Lewis (1950) gave Ss either 30 or 100 trials on a tracking task; Weitz and Adler (1971), using a foot-eye-hand coordination task, had "cycle time" as a criterion.

Mandler (1962) summarized the data concerning degree of original learning from the extensive verbal learning literature. First, he concluded that with small amounts of initial practice there is frequently a negative transfer effect, then a return to zero transfer with more practice, and finally, increasing positive transfer with even more practice. The best empirical generalization is that or a U-shaped function relating variations in degrees of original learning and amount of transfer. Second, he concluded that the degree of prior learning of an "incompatible response" leads to monotonically increasing negative transfer.

It is often difficult to argue with a U-shaped proposition; a given result can be interpreted as being on one side of the "U" or the other, depending upon whether transfer was positive or negative. In general, however, the results of the 13 studies reviewed are basically in accord with Mandler's
The second conclusion is that excessive practice with "wrong" responses is detrimental to transfer.

In summary, "amount of practice" undeniably has an effect on transfer; unfortunately, the effect has been demonstrated to be either positive or negative, depending on "other" considerations. These "other" considerations are not clearly specifiable from the literature; perhaps the only worthwhile generalization from this set of studies is that too much practice on inappropriate responses is bad.

**Previous Experience.** These five studies, as opposed to studies of "degree of original learning," dealt with "kind of original learning." Specifically, these studies were designed to test whether general or specific training was more beneficial for transfer. This problem is often stated in terms of the utility of "strategy" training, and stems from mediation theories of transfer. In most cases, Postman's (1962) 2-process theory (specific and non-specific transfer) is taken as a point of departure. As was true for the "Amount of Practice" variable, most of the work in this area has been done within verbal learning paradigms. Neiber (1968) states the general issue:

"Recently, Postman (1962) has hypothesized that, in the case of verbal learning, the tendency for negative transfer to first increase then decrease as first-task mastery increases results from the simultaneous operation of two processes: specific transfer (response competition), and non-specific transfer (warm-ups and learning-to-learn). Both processes gain strength as first-task practice increases, but non-specific transfer is thought to approach asymptote late." (p. 209).

The hypothesis is that this slowness in non-specific transfer causes the initial negative transfer. If through some sort of "non-specific pre-training," this initial decrement could be made up, the negative transfer could be reduced or eliminated. What is considered as non-specific training varies across experiments; in the documents which were reviewed, this ranged from amount of irrelevant cues in training (Overing and Travers, 1967) to number of years as a test pilot (Geiselhart, 1966).

Stated in more general terms, the issue of what kinds of tasks should be used to improve transfer is of utmost importance for the design of
training devices. The same general issue appears under other variable headings (e.g., part-whole training, adaptive training, stimulus pre-differentiation, stimulus variability). Restating the same issue still another way, the concern is how to train for transfer. Contributing evidence from the present set of studies is minor; the issues seem to revolve around the detail of specificity of the operational task and the conditions of performance of the operational task. If the degree of specificity is low and the task will be performed in various conditions, it appears that some initial, non-specific or generalized training may be appropriate. On the other hand, if the operational task is well specified (in the sense that the task is repetitive) and performed under uniform conditions, specific training might be the better technique.

Trainee (Student) Characteristics. This variable is usually labeled "individual differences" or "student aptitude." While only eleven studies reviewed in this survey dealt explicitly with individual differences, the more general literature in this area is enormous. For the most part, this variable has been considered as "given"; individual differences along any of several dimensions are assumed, and training programs have been designed to fit the maximum number of trainees. The only dimension of individual difference that has been systematically investigated is the AFQT (Armed Forces Qualification Test). For example, Fox, Taylor, & Caylor (1969) divided trainees into high (90-99), middle (45-55), and low (10-21) AFQT scores. These trainees were tested on a variety of military training tasks. Not surprisingly, AFQT score was a potent variable across all tasks (except for simple reaction time). Other studies explored relationships between individual difference data and "learning style" or "training methods."

One important outcome of these studies is the general postulate that it is often necessary to have some idea of what trainees "know" (in terms of specific skills and knowledge) in order to maximize device effectiveness. The critical issue is to delimit the dimensions of individual differences which may serve as a basis for selection or which should be considered during device design. These conclusions need further systematic investigation.
Training Requirements. The five studies subsumed under this label illustrate the point made above. Different specifications of operational task demands and training requirements should result in different methods of training. For example, Ward and Senders (1966) had all subjects performing a compensatory tracking task; however, different groups were given different instructions as to what was desired. The groups were transferred to the same tracking task but with different feedback conditions than they were trained on. Group differences on the transfer task were a function of the discrepancy between the instructions in the training task and the displayed feedback information in the transfer task.

These studies again point out the need for a clear specification of training objectives and operational tasks. This is important to make sure that what is being trained is truly what will be performed in the operational setting.

Adaptive Training, Stimulus Predifferentiation, and Part-Whole Training. The three studies dealing with Adaptive Training, the six with Whole-Part Training, and the eleven with Stimulus Predifferentiation all incorporate the same basic idea: training and transfer can be improved through the use of techniques which teach selected components of the total task. "Part-Whole" training is the most general concept; any component of the total task, when given special treatment, constitutes part training. "Stimulus Predifferentiation" refers to the part-training technique where the stimulus elements, typically those difficult to discriminate, are given special training. "Adaptive Training' is a more modern term for part-training where the "parts" chosen and/or the advancement of learning are a function of the trainee's performance.

There is insufficient evidence on adaptive training and stimulus predifferentiation to form any general conclusions. Common sense is probably the best guide; where the stimuli are difficult to discriminate, stimulus predifferentiation is a useful tool. Similarly, where it can be shown that some parameter of a complex activity is not only difficult to master but also capable of independent manipulation, that parameter could be made "adaptive." (Incidentally, the term "adaptive parameter" is often misused. Technically, an adaptive parameter is part of a closed-loop system;
some output transform is fed back to the operator, who then adjusts his performance accordingly. Simply sequencing problem difficulty from easy to difficult is not "adaptive" training; some variable must be altered as a function of output to be adaptive.) In order to best use adaptive training, the literature suggests that the selected variable must be central for learning. It is difficult to clarify this statement, other than to say that often some non-crucial parameter is selected; when this occurs, trainees become dependent on irrelevant feedback.

The issue of part- vs. whole-training is a bit more complicated. There are two potential benefits associated with part training as opposed to whole training:

1. Equal or lesser amounts of part-task practice could yield equal or higher levels of performance than whole-task practice.
2. To accomplish the same goals, the cost and complexity of simplified equipment for part-task training will be less.

Three kinds of tasks have been used in studies of part-whole learning:

1. complex sequential tasks,
2. tasks where parts must be time-shared, and
3. verbal learning tasks.

For sequential tasks, there are conflicting data, but the whole method is slightly better if the sequence is long and complex. Practically, however, when the "amount" of material is small, there may literally be no difference between part and whole practice; when it is large, part learning is a practical necessity. In time-sharing tasks, part-training should be used with caution and should be followed with a nominal amount of whole task practice. Investigations of part- versus whole-training in verbal learning studies have yielded mixed results since type of training has usually been confounded with "list differentiation" or "list organization."

Augmented Feedback. There were fifteen studies reviewed that dealt with augmented feedback. In general, the issues here are similar to
the issues stated above under "adaptive training." The system or device designer must be sure not to force trainees to attend to information that can be used as a crutch; this artificially improves performance during training but is deleterious when removed in the transfer situation.

"There is a cost associated with each class of information processed by a human operator. There is an element of workload associated with each element of displayed information, and that the magnitude of this workload is a function of the relative redundancy of the information displayed" (Ward and Senders, 1966).

Restating the last point, augmented feedback should be correlated in some way with the "natural" feedback. In addition, as a general guideline, augmented feedback should be "faded out" as practice progresses. (Further elaboration of this variable may be found later in the report).

Knowledge of Results (KOR). Only six studies dealt explicitly with knowledge of results. However, the literature for all types of tasks concerning KOR is quite large. Adams (1971) reviews a large body of data concerning KOR in simple motor tasks, and several books have been written on the topic. Historically, KOR was one of the first variables systematically investigated by experimental psychologists.

The most general conclusion from the present review is that some form of KOR is vital for learning and transfer. Furthermore, the KOR must be relevant for the trainee and appropriate for his level of ability. These two seemingly vacuous statements mean that a trainee needs some information as to what he is doing, and that this information must be in a form that he can understand and make use of.

CATEGORY II - DEVICE VARIABLES

Fidelity (Similarity). Probably no other issue in transfer of training and training device effectiveness has generated as much interest as the question of fidelity. Thirty-seven studies in this review dealt with some aspect of fidelity, either environmental, stimulus, or response. Fidelity has been the cornerstone of several device effectiveness models and the basis of most transfer of training theories. Many of the theoretical
arguments for considering fidelity are given in the modeis and theories sections above, and these will not be repeated here. As indicated above "fidelity" technically refers to the physical similarity, in an engineering sense, between two devices. However, common usage has made the terms "fidelity" and "similarity" virtually identical; people refer to degrees of fidelity just as they refer to degrees of similarity.

Given the theoretical and practical importance of fidelity, it is hard to understand why the data are so unconvincing in support of fidelity as a direct predictor of training effectiveness. For example, a number of authors are willing to accept the central role played by fidelity or similarity in positive transfer:

"The data regarding stimulus similarity and transfer tend to be quite consistent. In general, where stimuli are varied and the responses are kept identical, positive transfer increases with increasing stimulus similarity. This generalization has been supported by investigations using a wide range of learning tasks such as verbal paired associates, paired associates using visual forms as stimuli, expectancy learning, motor skills, and mediation tasks" (Ellis, 1969, p. 402).

"Concerning the factor of stimulus similarity there has never been any serious disagreements of experimental evidence with the following rule: Positive transfer increases with the degree of similarity of the stimuli of the initially learned task to the final task" (Gagne, 1954).

"In general, it can be stated that the amount of transfer expected to occur in flight simulator application seems to be proportional to the degree of fidelity provided. Although part-task simulators are usually cheaper and lower fidelity than whole-task simulators, they can be very useful for the learning of specific tasks. However, their shortcomings can be traced back to the lack of fidelity, particularly motion simulation" (Gerathewohl, 1969).

However, high fidelity does not always increase or insure positive transfer:

"The findings . . . seriously question the assumption that the best simulator for training purposes is the one that bears the closest physical resemblance
to the aircraft. Under rigorously controlled experimental procedures and using pilot performance as the primary criterion measure, both jet and non-jet pilots transferred successfully to full simulation conditions after training on greatly reduced conditions. Moreover, the pilots were unable to discriminate between the varying simulation conditions (Voss, 1969).

Similarly, other authors have commented that it is the "psychological" or "operational" realism which determines transfer and not the physical similarity (fidelity) of the devices. The empirical data are as varied as the above quotations: high fidelity improves transfer (e.g., Hammerton and Tickner, 1967); has no effect (e.g., Cox, et al., 1965); or produces negative transfer (in the A-B, A-B_1 verbal learning paradigm; see Ellis, 1969). It is clear that the effects of fidelity on transfer are large; however, these effects are modified (or determined) by "other" considerations. "Type of task" is certainly one of these, as is "task difficulty," "amount of practice," and so on.

The general conclusions are that: 1) some measure of fidelity could be used to estimate the "degree of representativeness" of a task in training; 2) the question of how well a task is represented in a training device has some a priori value; 3) some data do exist which suggest that the effect of fidelity varies as a function of type of task; and 4) fidelity per se is not sufficient to predict device effectiveness, mainly due to the scaling problem in non-laboratory tasks. Perhaps on a gross level some measure of "task communality" might be feasible as a parameter of similarity in a model of training device effectiveness. In tasks where subcomponents are well-specified, a more detailed measure of degree of fidelity might be a useful predictor of transfer.

Motion Simulation, Stimulus-Response Associations, Control Parameters, Display-Control Relationships. Several other variables in the Device category dealt with specific aspects of the fidelity issue. Motion Simulation, a sub-topic of environmental or stimulus fidelity, deals with the question of whether motion cues should be added to simulators. The finding is that the addition of these cues certainly changes the subject's
behavior and increases fidelity; whether this is good or bad for training and transfer is still open to question. **Stimulus-Response Associations** deals with two related issues: 1) is the "fidelity" of S-R bonds important for transfer; and 2) can the number of S-R bonds in common, or the number of bonds reversed, serve as a predictor of transfer? **Display-Control Relationships** (Compatibility) contains studies unique in that the **interface** between stimuli and responses in the training and transfer tasks was altered in some way. If the relationship was changed (from Task 1 to Task 2) or the relationship was different from population expectancies, marked effects on transfer were produced.

**Control Parameters** looked at variations in some controlled element (e.g., transfer from a velocity- to an acceleration-tracking task), or in variations of some vehicle dynamics from training to transfer tasks. Since, in these studies, it was difficult to forecast the effects of these variations on response fidelity or similarity, they were given a separate listing.

**Device Characteristics and Utilization.** This set consisted of fourteen studies that examined issues relating to: when in the training program a device should be used, how much time should be spent in simulators, and some specific device evaluations. Since each device was different, communalities and generalities among these studies were non-existent.

**CATEGORY III - TASK VARIABLES**

**Task Difficulty.** A relatively large number of studies (14) have dealt with the issue of whether training on an easy or a more difficult task (relative to the transfer task) results in greater transfer. A number of studies, summarized by Holding (1962), have shown that transfer may be considerably affected by differences in difficulty between tasks, in a manner which results in asymmetrical transfer relationships. That is, transfer may be greater in one direction than in the reverse direction between two tasks which differ in difficulty.

The asymmetrical effect cannot result from a similarity relationship, since if task A is similar to task B is by definition equally similar to task A. Studies of the have sometimes shown better difficult-to-easy transfer, and sometimes better easy-to-difficult transfer.
Unfortunately, there is no clear, single difference between the tasks which gives rise to these opposed effects.

Some progress may be made by considering two opposing tendencies. There is one feature of task relationships, often associated with differences in relative difficulty, which favors asymmetrical transfer in the difficult-easy direction. Many pairs of tasks are related by inclusion. In general, there will be greater transfer from the including to the included task than in the reverse direction; the included task will have been learned during practice on the including task, whereas practice on the narrower, included task will not involve learning of the additional components of the broader task. Thus, transfer will be greater from general radio servicing to output amplifier servicing, or from moving vehicle firing to static range firing, or from general tank handling to practicing gear shifting, because in each case the first task includes all the components of the second, while the reverse is not true.

Of course, this advantage will only hold if the first task is adequately learned. In some cases, the inclusive task may be so difficult that no learning takes place, so that the difficult-easy outcome is reversed. This appears to be the case in some perceptual tasks, where it appears that a difficult discrimination cannot be learned at all until an easier one is mastered (Day, 1956). The sense in which one discrimination "includes" another illustrates a useful extension of the notion of inclusion. For example, it is possible to describe the skill involved in shooting at a 5° target as (psychologically) including the skill in hitting a 10° target. As before, transfer will favor the difficult-easy direction.

On the other hand there is a group of factors, which may be called performance standards, which favor the easy-to-difficult direction of transfer. If a task is not only easier, but usefully easier, a trainee may learn efficient working methods, or habits of accuracy, or insights into the structure of the task, which he cannot acquire in a more difficult version. For example, trainees learning an intermediate level tracking task did better when transferred from an easier, automatically guided version than when transferred from difficult, normal practice in tracking a complex course (Macrae and Holding, 1966). The explanation appeared
to be that the guidance made the task usefully easier by selectively reducing the overload imposed by the complex task. Relieving the task of motor demands allowed the trainee to develop strategies of perceptual anticipation which held good on transfer to the new version. In this and similar cases, the trainee would also be expected to develop high standards of accuracy during the easy preliminary task which would be carried over to the difficult transfer task.

The outcome of any asymmetrical transfer comparison will therefore depend upon the delicate balance between difficulty-favoring inclusion factors and easiness-favoring performance efficiency factors. However, although inclusion will be associated with relative difficulty over most of the range of task difficulties, the performance standards effect will depend crucially upon absolute difficulty. When the difficult task is very difficult, it will no longer be possible to carry the performance standards over from the easier version, and the inclusion effect will supervene. Thus, even when an easy mean task level produces easy-to-difficult superiority, increasing the difficulty of both tasks may reverse the effect (Holding, 1962).

Task Duration, Task Organization, Stimulus Variability. These three variable headings deal with various aspects of task dimensions. Task Organization (5 studies) is a term taken from verbal learning studies; it refers, in general, to some scale of meaningfulness or "integrity" of the task or sub-task. Task Duration (one study) refers to the possible effect of length of the task on training and transfer. Stimulus Variability (six studies) refers to situations where the critical stimuli in the tasks are either exemplars of an extremely large set (e.g., enemy vehicles) and/or may occur under perceptually confusing circumstances. Due to this property, it is often necessary to provide exposure to an array of stimuli and to a range of background conditions.

Generalizations from these data sets are unwarranted due to the small number of studies and the lack of common features of the studies reviewed.

Task Analysis. These eleven studies dealt with different approaches to task analysis; some cited empirical "tryouts" of existing methods, while others discussed training implications of these methods. This general topic is discussed more fully above.
While the number of studies directly comparing types of tasks is small, the weight of evidence justifies the use of type-of-task as a dimension or axis of any model of training device effectiveness. Stolurow (1964) points out that the great amount of information on learning and transfer can be used by training analysts only to the extent that the generalizability of the findings across tasks can be specified. Thus he suggests that empirical findings about learning and transfer must be systematically related to a task taxonomy so that any moderating effects of task on the empirical findings can be noted and used by the training analyst. In a series of experiments he demonstrates that task categories do interact with various empirical training effects. Fitts (1962) puts it this way:

"The importance of an adequate taxonomy for skilled tasks is widely recognized in all areas of psychological theorizing today. A taxonomy should identify important correlates of learning rate, performance level, and individual differences" (p. 178).

Thus, findings for most of the variables in the "training-variable" and "device-variable" categories should be viewed with "type of task" as a modifier.

SUMMARY AND IMPLICATIONS: TRAINING PRINCIPLES

Rationale. The original purpose in reviewing the substantive issues was to decide which variables clearly influenced transfer of training. However, given the fragmented nature of the literature and the difficulty in presenting results in summary form, it became apparent that simply specifying variables known to influence transfer would be of limited value for the development of a model of training device effectiveness. Consequently, attention turned to the possible use of these variables as inputs to a set of generalizable "principles" bearing on a training device or on the more general training system. The rationale for attempting to synthesize principles of training needs amplification; for, in essence, the answer to the question of how this information would be used directly implies a model for training effectiveness evaluation.
Consider, as a hypothetical example, the following situation: Three devices, A, B, and C, are proposed as trainers for a complex perceptual-motor skill. The three devices are identical, except that A is equipped with a mechanism which provides knowledge of results (KOR) in the form of augmented feedback, B provides KOR based on natural feedback, and C, for some reason, provides no KOR. The problem is to determine which device should be selected. The literature reveals what seems to be directly relevant information. Lucas, Heimstra, and Spiegel (1973) compared the performance of two groups on a driving task. One group was provided with knowledge of results during training, and the other group was not. The knowledge-of-results group proved to be far superior on the transfer task. Thus, KOR was demonstrated to have a direct and positive effect on transfer. We can generalize from the Lucas et al. study to the three devices and, on that basis, can rule out device C. Before deciding between devices A and B, however, more information is required. For instance, is augmented feedback better than natural feedback? Are we willing to assume that the effect of augmented feedback is independent of the task? Are all forms of augmented feedback equivalently good? Does the level of skill of the trainee, amount of practice, or difficulty of the task interact with the effect of augmented feedback? For instance, Ward and Senders (1966) studied augmented feedback in a tracking task. They found that:

"... supplemental information required attention of the tracker and interfered with his primary task performance. Apparently any added non-redundant visual signal interferes with the performance on the main task."

In other words, they found conditions where augmented feedback during training produced negative transfer when the feedback in the transfer task was not correlated with the augmented feedback in the training tasks. Salvendy

2 "Augmented feedback" is not synonymous with "knowledge of results." Augmented feedback refers to information displayed to the trainee that is not inherent in the task. Knowledge of results could be provided by augmented feedback or it could be directly sensed by the trainee as a result of his actions.
and Harris (1973) used 18 types of augmented feedback in a psychomotor skill. The different types were not equivalent, with some producing positive transfer, others no transfer, and still others negative transfer.

The implication of the above example is that the probability is very small of finding the specific study that can provide sufficient information to decide between devices A and B. However, collectively, the set of studies on augmented feedback and KOR might be able to provide enough generality to enable the choice to be made. Furthermore, synthesis of the literature might specify the limits of the Ward and Senders finding and suggest an implementation of the augmented feedback. For example, Welford (1968), in summarizing a large body of data from motor tasks, proposes the following "principle":

"A subject must have some cues to the results of his actions if he is to perform accurately at all, and training procedures will be effective insofar as they help him to observe and use such cues as are inherent in the task for which he is being trained. They will fail insofar as they provide him with extra cues on which he comes to rely but which are not available when he changes from training to the actual job."

This "principle" is both a summary, in the form of an empirical generalization, and a guide for the evaluation of a training device. Furthermore, it is consistent with the specific experiments mentioned above and consistent with more general "principles" of training (see, for example, Willis and Peterson, 1961). Thus, a synthesis of the literature should make use of both the direct empirical results and a set of integrated principles of training derived from those data. The principles would be used to specify further device dimensions of importance. For example, Adams (1971) states the following principle: "In simple motor tasks, withdrawal of knowledge of results produces deterioration of performance when level of training is low or moderate." This suggests that the evaluator must pay attention to the type of task and level of training, in addition to features of the KOR itself.

Having demonstrated the utility potentially inherent in training principles, concern can shift to their availability. The situation is relatively
good since, as mentioned earlier, several methodists have attempted to use or synthesize principles. These principles have been derived from a number of sources and can be evaluated as to their usefulness and importance for incorporation into a model. Ideally, we would like to "build" training principles from the empirical studies alone. Each principle would then be firmly based in data. However, since practical training knowledge probably exceeds its empirical and theoretical support, it would be unwise to ignore the "rules of thumb" or "common sense" of skilled training specialists. Likewise, it would be silly to ignore "principles" derived from learning theory. In effect, we are working "downward" as well as "upward" in generating training principles.

In the final portion of this section, a list of some of the more readily identifiable principles is presented. These statements are designed to serve as summaries of the relationships hypothesized to exist among variables within the "training-, device-, and task-variable" categories.

Training Principles. In reviewing the existing data, it became apparent that there were three major influences on how efficiently training takes place. These three influences encompass and account for many of the variables identified to date. The three major categories are:

1. Training management and control techniques
2. Conditions of first task performance
3. Conditions of feedback and knowledge of results

Training management and control techniques, in this context, refer to the fact that the learning situation (e.g., training device) must provide some type of guidance to the student or trainee in order to promote efficient learning. The training process must be orderly, with sufficient structure to provide for sequencing of difficulty, monitoring of progress, and demonstration if required. Increasing the level of problem difficulty until it approaches that found in the operational situation, whole versus part training, and stimulus predifferentiation are all techniques which would fall under this heading.

The conditions of first task performance category focuses on provision of an opportunity for practice and includes all of the identifiable
variables associated with it. The role of active practice, spacing of practice, and amount of practice all can affect the efficiency of learning, retention, and transfer.

The third factor which must be present in some form is a system for feedback and correction. Again, such variables as the timing of feedback, its precision, what occurs in the intervals between feedback, and whether or not augmented feedback is used, all affect training efficiency.

For efficient training, all three of these factors must be present in some form. Principles comprising each category are presented below.

**Training Management and Control Techniques**

1. Ensure that relevant subordinate capabilities have been thoroughly learned before calling on transfer to aid the learning of "advanced" capabilities.

2. Transfer is enhanced by the variety of previous knowledge.

3. With very complex tasks, instruction in principles yields better results than laying down a detailed drill, while with simpler tasks the drill is at least equally effective.

4. Where the whole task is a closely coordinated activity such as aiming a rifle or simulated flying of an aircraft, it is better to tackle the task as a whole. Any attempt to divide it up tends to destroy the proper coordination of action and subordination of individual actions to the requirements of the whole, and thus outweighs any advantage there might be in mastering different portions of the task separately.

5. Where the task involves a series of component actions which have to be performed in the correct order but each is largely independent of the others, there seem to be advantages in practicing the different components separately.

6. If errors could be prevented in the first few trials (e.g., guidance), mastery of the task should be very much quicker.
7. Guidance during training is beneficial when tracking movements have to be made with an incompatible control-display relationship.

8. Guidance does not aid simple repetitive movements, but aids learning complex courses.

9. If two or more tasks have to be learned, it is most beneficial to begin with the one which elicits the greatest care and effort towards the attainment of a high standard of performance. However, if the subject (S) was not allowed to continue to practice the more difficult task until a point of reasonable mastery, he would be left with an inadequate comprehension of the task, and transfer to a simpler task might be confused and less satisfactory than if he had tackled the easier task first.

10. The more sub-tasks there are in the overall task, and the more they interact with one another, the more opportunity there will be for improvement, and therefore the longer improvement will continue.

**Conditions of First Task Performance**

1. The usefulness for transfer of any learned capability will be increased if it is practiced in as wide a variety of situations as possible.

2. Continuous practice facilitates mastery of complex, meaningful material and the establishment of coordinated rhythmic activity.

3. Continuous practice seems to be preferred by older trainees.

4. Spaced practice is more efficient than continuous if only the actual duration of the sessions is counted and the time between sessions is ignored. When the time between sessions is included, continuous practice is usually more efficient.

5. Effectiveness of spacing practice depends on what is done during the times between practice periods: (a) If they are spent in rehearsal of the material, learning will benefit, unless the task is fatiguing, in which case continued practice may depress subsequent performance. (b) If time between practice periods are spent on another task,
learning or later recall of the first task may be impaired, the degree of impairment depending on the degree of similarity between the two tasks.

6. Very brief pauses between practice sessions should be as effective as longer ones.

7. "Mental practice," in which the S performs a task in imagination, can often be substituted for a substantial amount of practice involving full performance with little, if any, loss of effectiveness.

8. Relatively little learning occurs if Ss are passive spectators or even passive performers, but that they must be involved in active decisions and choices about what they are doing, and it is these that they will retain, whether they are right or wrong.

**Conditions of Feedback and Knowledge of Results**

1. Performance improvement in acquisition depends on knowledge of results (KOR). The rate of improvement depends upon the precision of KOR.

2. Delay of KOR has little or no effect on acquisition.

3. Increasing the post-KOR interval up to a point will improve performance level in acquisition.

4. The type of activity in the KOR delay or post-KOR delay interval does not influence acquisition.

5. Withdrawal of KOR produces deterioration of performance when level of training is low or moderate.

6. When KOR is delayed in acquisition, and S engages in deliberate verbal or motor activity during the delay interval, the affect of KOR withdrawal is poorer performance than when S rests.

7. When KOR is delayed in acquisition, and S rests during the delay interval, the affect on performance when KOR is withdrawn is no different than when immediate KOR is used.

8. Activity in the post-KOR delay interval during acquisition worsens performance when KOR is withdrawn.
9. After a relatively large amount of training, learning can continue when KOR is withdrawn.

10. Pre-training methods need to take care not to make the $S$ dependent upon the extra cues (augmented feedback) provided in the early stages of training and, thus, to hinder the changeover to more direct relations between input and output at a later stage.

11. Information about the correctness of action should be available quickly.

12. The manner of conveying KOR is important:
   a. Effectiveness is greatest when the information is clearly and simply related to the action performed. Any distortion or equivocation in the information fed back to the $S$ will reduce its effectiveness.
   b. Unduly full or complex information may be partly ignored or may confuse the $S$.
   c. The information given should indicate the discrepancy between what is required and what has been achieved, rather than merely give a reminder of requirements or some broad measure of achievement.

13. $S$ must have some cues to the results of his actions if he is to perform accurately at all, and training procedures will be effective insofar as they help him to observe and use such cues as are inherent in the task for which he is being trained. They will fail insofar as they provide him with extra cues on which he comes to rely but which are not available when he changes from training to the actual job.

14. KOR acts as an incentive; it can be intrinsically motivating.

The principles stated above should be viewed as neither complete nor definitive. It is anticipated that they will be modified and it is likely that the list will be expanded. In order to be truly useful, the data and information must be organized to facilitate the evaluation of a given training device or training device concept. In the following section, the rudimentary articulation of a structure is provided in the form of a preliminary model for predicting training effectiveness.
PRELIMINARY MODEL.

In model development, the goal is not only to identify the relevant factors, but also to integrate them into meaningful relationships, preferably in a quantitative manner. Historically, this has meant reducing very complex relationships down to basic concepts and generalizations. The same approach was used in constructing the current model. The basic ingredients were the data and principles identified from the exhaustive search of the relevant literature. This literature review did produce a few generalizable principles and suggested some organizational structure for assimilating and interpreting the data. Unfortunately, however, the literature consists mostly of studies with such different approaches, measures, controls, and variables that it is impossible to reconcile, let alone assimilate, all of their findings. Despite the long-standing interest in learning, retention, and transfer, there is little which can be systematized into coherent principles or laws.

Deficiencies in the existing literature, however, should not stop attempts at systematization. Where data are available, they should be used. Where they are not available, or are questionable for one reason or another, certain assumptions need to be made. These assumptions should be documented so that as additional data become available, they can also be incorporated into the model. The model itself should, in fact, become the basis for a research program directed at acquiring the necessary data.

The overwhelming implication from the literature reviewed to date is that the effectiveness of a training device depends upon two major classes of variables:

1. Those associated with developing a training device which does, in fact, elicit the behaviors which are required in the real or operational situation. (These have been called "Device Variables" above.)

2. Those associated with actually learning these behaviors. (These have been referred to above as "Training Variables.")
A distinction is drawn here between these two sets of variables because it is necessary at the outset to recognize the dual nature of the requirements for a training device. It must first and obviously in some way elicit or require that the appropriate behaviors occur. What the trainee does on the device should prepare him to perform those tasks which will be required of him in the operational setting. More simply stated, the training device should be a valid instrument. A training device is valid to the extent that it trains the student on the tasks he is supposed to be trained in. Validity, therefore, refers to the Appropriateness of the training.

Appropriateness of the training device is of key importance, but is not the only factor that must be dealt with. In addition, the Efficiency of the training device needs to be considered. Learning can take place under many kinds of conditions. "Incidental learning" occurs with little motivation or awareness. However, the efficiency (more speed, less cost or effort, greater retention, etc.) with which learning takes place can be greatly affected by the conditions under which it occurs. Basic and applied research in learning, training, and education have supplied us with data on some of these variables and their effects.

In considering Efficiency, an additional and important point should be made. Transfer is not only produced by the variables that affect it directly; acquisition and retention are also highly relevant issues. In fact, transfer cannot take place unless acquisition and retention do. Therefore, the variables affecting acquisition and retention also apply to transfer. This basic assumption is the underlying foundation for the entire portion of the model dealing with Efficiency, which has been broadened to mean efficiency in learning and retention as well as transfer.

Thus, a training device must train the appropriate behaviors and do so efficiently. Its effectiveness is dependent upon both considerations— Appropriateness and Efficiency.

The major conclusion that was drawn from the review of existing models and methods of training device evaluation was that all previous models ignore at least one of the key issues involved in predicting transfer.
<table>
<thead>
<tr>
<th>EFFICIENCY</th>
<th>TRAINING PRINCIPLES &amp; TECHNIQUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of Training</td>
<td>p_1, p_2, \ldots, p_x</td>
</tr>
<tr>
<td>Learning Difficulty</td>
<td></td>
</tr>
<tr>
<td>Repertory</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>APPROPRIATENESS</th>
<th>TRANSFER POTENTIAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Similarity</td>
<td></td>
</tr>
<tr>
<td>Criticality</td>
<td></td>
</tr>
<tr>
<td>Communality</td>
<td></td>
</tr>
<tr>
<td>Behavioral Categories</td>
<td>c_1, c_2, c_3, \ldots, c_k</td>
</tr>
<tr>
<td>Task Description</td>
<td>t_1, t_2, t_3, \ldots, t_n</td>
</tr>
</tbody>
</table>

Figure 9. Preliminary structural model.
Several critical issues were identified that must be resolved in order for a valid prediction to be made. They are:

**Appropriateness**
1. A determination of what (skills and knowledge) the device is supposed to be training.
2. A determination of the tasks and equipment which are in common between the training situation and the operational situation.
3. An assessment of the importance of the tasks represented in the training situation to operational performance.
4. An evaluation of how well the device represents these critical tasks.

**Efficiency**
1. An analysis of the type and amount of learning which must be accomplished in the training situation (i.e., what the trainee knows versus what he must learn).
2. A determination of the manner in which the training device (including the training system in which it is embedded) proposes to make up this learning deficit.

The tentative structural model (Figure 9) represents a framework within which these issues are made explicit. This model is a synthesis of the previous efforts, theoretical analyses, and empirical data reviewed in the preceding sections. It is basically a training content by training process model, in accordance with the conclusions of this section. A matrix representation of the model is, in our opinion, the most feasible way of addressing the issues cited above. The implications of these issues for the model are considered sequentially.

First, what is the device supposed to be training? The only reasonable way to answer this question is through some sort of task analysis of the training requirements and/or the operational situation. It is not sufficient to say that "this device is to be used to teach gunnery"; we must know what the gunner's job consists of before training effectiveness can be

---

3The "efficiency" aspect of training effectiveness will be moderated by a host of potent variables external to the device itself. These variables (e.g., device acceptance, other instruction, etc.) lie outside the purview of the current device effectiveness model, but would be incorporated in a training system model.
evaluated. One orthogonal dimension of the model, labeled "Training Content," represents two levels at which this question can be answered. One possibility is a detailed task description: an enumeration of everything the operator does in the real situation. Another possibility is a task categorization; this represents the classification of task components into behavioral categories. In a sense, a task categorization "reduces" the task description into another sort of useful information.

The next three issues center around the preceding task analysis. Collectively, they can be summarized as one more general question: Independent of training techniques (and other moderator variables), what is the transfer potential of the device? If we assume that the trainee becomes proficient on the tasks present in the device, will he then meet the training requirements? In order to answer this question, the three remaining appropriateness issues must be addressed.

If we assume that the trainee will become proficient on the training device, it is evident that the device should incorporate as much of the operational task as possible. This is represented in the model as "Communality." However, the notion of communality is modified by the importance of the represented tasks to operational performance. If some task is absolutely vital to operational performance, it should be represented in the training device in order for the device to be a potentially effective trainer. This issue is represented in the model as "Criticality." Finally, the transfer potential is influenced by how well the device represents these critical tasks. This issue acknowledges the inescapable conclusion that similarity or fidelity plays a role in transfer. We must know, at some level of description, how well the tasks incorporated in the device represent those same tasks in the operational setting. Although actual transfer is some (as yet unknown) function of similarity (and the interactions between similarity and other variables), transfer potential must vary directly as a function of similarity. Some measure of task correspondence is represented as "Similarity" in the model.

The next issues deal with the second major consideration -- efficiency. A problem in device design and evaluation which is often ignored is simply
how much trainees actually have to learn. A device can incorporate all tasks of the operational situation, and be an absolutely faithful representation of the real world, but if the skills involved are already possessed by the trainee, the device probably is of little value. The "Learning Deficit" portion of the model addresses this problem in three different ways by: (1) determining whether the skill and knowledge underlying the task are already in the trainee's repertory; (2) estimating how difficult it is to learn the task; and (3) establishing the proficiency requirements for the training to be accomplished on the device.

Up to this point in the model, the evaluator did not have to know anything about "how to train." The next section of the model, "Training Principles and Techniques," deals with this issue. It represents an attempt to make direct use of the empirical data and principles developed in this section. It was implied there how these principles could be used. For example, in a given implementation of knowledge of results, the feedback which is given can be evaluated in terms of those principles concerning KOR and feedback. Furthermore, it might be possible to find relevant empirical data bearing directly on the particular implementation (i.e., either specific principles or previous experimentation for that type of KOR by task-type intersection). The procedure for this portion of the analysis has yet to be formulated. The guiding consideration in developing this portion of the model will be to make it explicit, so that the evaluator need not rely so extensively on "expertise".

Given the preliminary model in its present form, much remains to be done for its full articulation. Specific parameters within each of the major portions of the model must be specified. Data requirements and scaling procedures must be reconciled. Procedures must be refined for aggregating the information into a single predictive index. Subsequent reports will describe such refinements and will discuss results of validation research.
REFERENCES


Bertin, M. A., Colvin, R. P., Benfari, R., Lanchony, L., Logan, W., Metlay, W., Suwara, J., & Wallach, S. Experimental training situation analysis (TSA) and application of training analysis procedures (TAP) to two military systems. NAVTRADEVVCN 342-6, U.S. Naval Training Device Center, 1963, AD378180. (The report is confidential)


Geiselhart, R. The effect of prior experience on acquiring skill on a simulated inertial control task. Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio, 1966, AD641389.


Osering, R. L., & Travers, R. M. Variation in the amount of irrelevant cues in training and test conditions and the effect upon transfer. Journal of Educational Psychology, 1967, 58(1), 62-68.


Salvendy, G., & Harris, D. R. Effects of different modes of feedback on the acquisition and retention of psychomotor skills. Ergonomics, 1973, 16(3), 333.


Weitz, J., & Adler, S. The optimal use of simulation, New York University, Department of Psychology, 1971, AD730951.


BIBLIOGRAPHY


Adams, J. A., & Lewis, D. The evaluation of 'difficulty of task' under several different conditions of performance on the Modified Mashburn Apparatus. SDC57-2-8, Contract N5ori-57, State University of Iowa, 1949, AD639293.

Adams, J., & Lewis, D. The problem of controlling level of learning in studies of associative interference in psychomotor performance. SDC57-2-9, Contract N5ori-57, State University of Iowa, 1949, AD639294.


Andreas, B. G., Green, R. F., & Spragg, S. D. S. Facilitation effects between following and compensatory tracking tasks. SDC241-6-7, Contract N6onr-241, University of Rochester, 1952, AD126262.

Andreas, B. G., Green, R. F., & Spragg, S. D. S. Transfer effects in following tracking resulting from reversal of the display-control relationship on alternate blocks of trials. SPECDEVCE 241-6-10, Contract N6onr-241, University of Rochester, 1953, AD51494.

Andreas, B. G., Green, R. F., & Spragg, S. D. S. Transfer effects in compensatory tracking resulting from reversal of the display-control relationship on alternate blocks of trials. SPECDEVCE 241-6-21, Contract N6onr-241, University of Rochester, 1955, AD51494.


Barch, A. M. A demonstration of retroactive interference in pursuit rotor learning. SDC166-00-1, Contract Nonr-166(00), State University of Iowa, 1951, AD639306.


Bauer, M. Effects of absolute and conditional probability judgments on prediction in probabilistic inference. UMEA. University of Sweden, Department of Psychology, 1971, AD892109.


Bertin, M. A., Colvin, R. P., Benfari, R., Lanchany L., Logan, W., Metlay, W., Suwara, J., & Wallach, S. Experimental training situation analysis (TSA) and application of training analysis procedures (TAP) to two military systems. NAVTRADEVCEN 342-6, U.S. Naval Training Device Center, 1963, AD378180. (The report is confidential)


Boycan, G., & Kraemer, R. E. Training implications of the control and display data for the US/FRG MBT-70, M60A1E2, M551, and M60A1, Human Resources Research Office, Alexandria, Virginia, 1969, AD876285L.

Brecke, F., & Gerlach, V. Model and procedures for an objective maneuver analysis. Arizona State University, Instructional Resources Laboratory, 1972, AD758229.

Brecke, F., & Reiser, R. Critical components of flight instruction as perceived by instructor pilots and student pilots. Arizona State University, Tempe, Arizona, 1972, AD758277.

Brehmer, B. Concurrent learning of two single-cue probability learning tasks. UMEA. University of Sweden, Department of Psychology, 1971, AD891129.

Briggs, G. E. On the scheduling of training conditions for the acquisition and transfer of perceptual-motor skills. NAVTRADEVCEN 836-1, Contract N61339-836, Ohio State University, 1961, AD279206.


Briggs, G. E., Naylor, J. C., & Fuchs, A. H. Whole versus part training as a function of task dimensions. NAVTRADEVCEN 950-2, Contract N61339-950, Ohio State University, 1962, AD277617.

Briggs, G. E., & Wiener, E. L. Fidelity of stimulation. I. Time sharing requirements and control loading as factors in transfer of training. Ohio State University Research Foundation, Columbus Lab of Aviation Psychology, 1959, AD23100G.


Bryant, P. E. Comments on the design of developmental studies of cross-modal matchings and cross-modal transfer. *Cortex*, 1968, 4, 127-137.


Casperson, R. C., and Channell, R. C. Use of the operational flight trainer. NAVTRADEVCEN 1734-00-1, Contract Nonr-1734(00), Dunlap & Associates, Inc., 1957, AD643498.


Cohen, E. Is motion needed in flight simulators used for training? Human Factors, 1970, 12, 75-79.


Dougherty, D. J., Houston, R. C., & Nicklas, D. R. Transfer of training in flight procedures from selected ground training devices to the aircraft. Illinois University, Urbana Aviation Psychology Lab, 1957, AD149547.


Edenborough, R. A. User opinion of RAF flight simulators. RAF Institute of Aviation Medicine, Farnborough, England, 1968, AD857780.


Ellis, R. H. Task analysis reduction technique (TART) for the quantification of human performance. Naval Personnel and Training Research Laboratory, San Diego, California, 1970.


Garvill, J., & Molander, B. Asymmetric effects in cross-modal transfer. UMEA. University of Sweden, Department of Psychology, 1968, AD835431.


Gebhard, J. W. Psychological problems in coding information for visual displays. TR166-1-93, Contract N6onr-166, John Hopkins University, 1949, AD640077.

Geiselhart, R. The effect of prior experience on acquiring skill on a simulated inertial control task. Aerospace Medical Research Labs, Wright-Patterson Air Force Base, Ohio, 1966, AD641389.

George, C., & Miller, D. Performance of real and nominal crews under transfer conditions; two experiments. Texas Technical University, Lubbock, Texas, 1972, AD738176.

Gerall, A. A., & Green, R. F. The effect upon performance on a following tracking task of interchanging the direction of display movement controlled by each hand. SPECDEVCEN 241-6-24, Contract N6onr-241, University of Rochester, 1955, AD109355.


Gerlach, V., Brecke, F., Reiser, R., & Shipley, B. The generation of cues based on a maneuver analysis. Arizona State University, Instructional Resources Laboratory, 1972, AD758228.


Goldman, A. Training 90MM gun crews. Appendix V - Follow up study of recommended procedures and training aids. *NAVTRADEVEN 495-01-45, Contract Nonr-495(01), Ohio State University, 1956, AD642590*.


Green, R. F. The transfer of tracking skill as a function of task difficulty (target size). *SPECDEVCECN 241-6-13, Contract N6onr-241, University of Rochester, 1953, AD109345*.

Green, R. F. The effects on performance of a compensatory tracking task of interchanging the control of the axes of movement between the hands. *SPECDEVCECN 241-6-23, Contract N6onr-241, University of Rochester, 1955, AD110963*.


Hansen, D. N., & Dick, W. Memory factors in computer-controlled maintenance training. NAVTRADEVCEN 68-C-0071-1, Contract N61339-68-C-0071, Florida State University, 1969, AD697980.

Harby, S. F. Comparison of mental practice and physical practice in the learning of physical skills. SDC269-7-27, Contract N6onr-269, Pennsylvania State College, 1952, AD640878.

Harris, J. D. Habituationary response decrement in the intact organism. Psychological Bulletin, 1943, 40(6), 385-422.

Hedlund, J. L., & Lewis, D. Further attempts to demonstrate interference in the performance of rotary pursuit tasks. SDC938-1-3, Contract NNonr-93801, State University of Iowa, 1950. AD644634.


Higgins, N. C. Feedback in group instruction. Arizona State University, Tempe, 1972, AD754979.

Hirsch, R. S. Experimental evidence for improvements needed in rifle marksmanship training. SPECDEVCE 494-01-3, Contract Nonr-494(01), Tufts College, 1953, AD222931.


Horrocks, J. E., Bowlus, D. R., & Krug, R. G. Light anti-aircraft tracer observation and fire control with specific reference to the training problem. SPECDEVCEN 495-01-3, Contract 495(01), Ohio State University, 1952, AD642527.


Horrocks, J. E., & Fotheringham, W. C. Study of training requirements for tank tactical training. SPECDEVCEN 495-07-1, Contract Nonr 495(07), Ohio State University, 1954, AD642595.

Horrocks, J. E., Fotheringham, W. C., & Bowlus, D. Study of present status of training aids and devices in the Army field artillery training program. NAVTRADEVCEN 495-8-1, Contract Nonr-495(08), Ohio State University, 1956, AD642596.

Horrocks, J. E., Krug, R., & Bowlus, D. Training for antiaircraft artillery gunnery. SPECDEVCEN 495-01-1, Contract 495(01), Ohio State University, 1952, AD642505.


Kelly, J. C. Retention of voice communication training after one year. TR104-2-13, Contract N6ori-104, Purdue University, 1949, AD639085.


Kinkade, R. G. Augmented feedback and tracking skill. NAVTRADEVCEN 508-3, Contract N61339-508, Ohio State University, 1959, AD230999.

Klaus, D. J. The investigation of step size and error rate in programmed instruction. NAVTRADEVCEN 1208-1, Contract N61339-1208, American Institutes for Research, 1964, AD607782.


Lawrence, D. H. The evaluation of training and transfer programs in terms of efficiency measures. Stanford University, California, 1954, AD87172.


Lemke, E. A., & Randle, K. Effects of degree of initial acquisition, group size, and general mental ability on concept learning and transfer. *Journal of Educational Psychology*, 1969, 60(1), 75-78.


Levine, M., Morgan, R. L., & Neiberg, A. D. Influence of first task practice and intertask similarity on transfer of training in a symbol substitution task. Aerospace Medical Research Labs, Wright-Patterson Air Force Base, Ohio, 1964, AD608598.


Lewis, D., & Smith, P. N. Retroactive facilitation and interference in performance on the modified two-hand coordinator. SDC66-00-2, Contract Nonr-166(00), State University of Iowa, 1951, AD639362.


McAllister, D. E. The effects of level of learning and of overlearning on proactive and retroactive facilitation and interference. SDC938-1-8, Contract N90onr-93801, State University of Iowa, 1950, AD643413.

McAllister, D. E., & Lewis, D. Single-trial-per-task versus multiple-trials-per-task in the acquisition of skill in performing several similar tasks. SDC938-1-7, Contract N90onr-93801, State University of Iowa, 1950, AD643412.


Maccoby, N., & Sheffield, F. D. Theory and experimental research on the teaching of complex sequential procedures by alternate demonstration and practice. Boston University, Boston, 1957, AD842250.


Mahler, W. R., & Bennett, G. K. Special devices in primary flight training: Their training and selection value. SDC151-1-18, Contract N60r-151, Psychological Corporation, 1949, AD764033.

Mahler, W. R., & Bennett, G. K. Psychological studies of advanced Naval air training: Evaluation of operational flight trainers. SDC999-1-1, Contract N609r-999(01), Psychological Corporation, New York, 1950, AD643499.


Matheny, B. J. Two simulators for training pilots and controllers in air traffic control procedures. SDC71-16-15, Contract N60r-71(16), University of Illinois, 1955, AD70889.


Melton, A. W. Information encoding in training for efficient access in memory. Michigan University, Human Performance Center, Ann Arbor, 1973, AD758807.

Merjan, F. L. Sonar operator training: A review of the literature. NAVTRADEVCEN 322-1-1, United States Naval Training Device Center, 1957, AD141959. (This report is confidential)


Micheli, G. Training situation analysis report for the Marine tactical data system. NAVTRADEVCEN IH-18, U. S. Naval Training Device Center, 1964, AD376639. (The report is confidential)


Muller, D. Learning to learn as a factor in the transfer of stimulus predifferentiation. Psychonomic Science, 1970, 19(6), 317-318.


Murnin, J. A. Comparison of training media: Transfer of principles involved in a manipulative skill; operation of the aircraft load adjuster slide rule. SDC269-7-103, Contract N6onr-269, Pennsylvania State University, 1955, AD81663.

Naylor, J. C. Parameters affecting the relative efficiency of part and whole training methods: A review of the literature. NAVTRADEVCEN 950-1, Contract N61339-950, Ohio State University Research Foundation, Columbus, Ohio. Laboratory of Aviation Psychology, 1962, AD275921.


Overing, R. L., & Travers, R. M. Variation in the amount of irrelevant cues in training and test conditions and the effect upon transfer. *Journal of Educational Psychology*, 1967, 58(1), 62-68.


Parker, E. L. Application and design characteristics of generalized training devices. NAVTRADEVCEN 70-C-0309-1, Contract N61339-70-C-0309, Anacapa Sciences, Inc., 1971, AD733471.


Pomarolli, R. S. The effectiveness of the Naval Air Basic instrument trainer. Naval Aerospace Medical Institute, Pensacola, Florida, 1965, AD627218.


Puig, J. A. Motion in flight training: A human factors view. NAVTRADEVCEN IH-177, Naval Training Device Center, 1970, AD880445.


Quick, G. R., & Stark, E. Main Battle Tank - 70 (MBT-70) training device requirements study. NAVTRADEVCEN 68-C-0037-1, Contract N61339-68-C-0037, General Precision Systems, Inc., 1968, AD395927L. (This report is confidential.)


Roberts, K. C., & Taylor, C. L. The instructional cue. Arizona State University, Instructional Resources Laboratory, 1972, AD754229.


Salvendy, G., & Harris, D. R. Effects of different modes of feedback on the acquisition and retention of psychomotor skills. *Ergonomics*, 1973, 16(3), 333.

Sandberg, K. O. W., & Lipschultz, H. L. A survey of the importance and use of controls and displays on radar console panels. TR166-1-7, Contract N50ri-166, Johns Hopkins University, 1947, AD639376.


Shephard, H. Losses of skill in performing the standard Mashburn task arising from different levels of learning on the reversed task. SDC938-1-9, Contract N9onr-93801, State University of Iowa, 1950, AD643405.


Smith, P. N., & Lewis, D. Simultaneous functioning of retroactive facilitation and interference. SDC166-00-3, Contract Nonr-561(00), State University of Iowa, 1952, AD128719.


Spieth, W., & Lewis, D. The effects of alternating practice on the performance of two antagonistic motor tasks. SDC938-1-6, Contract N900r-93801, State University of Iowa, 1950, AD643411.


Trow, W. C. The problem of transfer: Then and now. In H. F. Clarizio, R. C. Craig, & W. A. Mehrens (Eds.), Contemporary Issues in Educational Psychology, (See 44:8) 279-285.


Tufts College & Documentation, Inc. Coordinate index and abstracts of training device literature. NAVEXOS P-1735, Contracts Monr-1929(00) & N61339-218, 1957, AD218923.


U.S. Continental Army Command Headquarters. Training systems engineering of training (course design). Fort Monroe, Virginia, 1972 (Con Reg 350-100-1).


Viteles, M. S. The aircraft pilot: Five years of research, a summary of outcomes. Psychological Bulletin, 1945, 42(8), 489-526.


Vris, T. A comparison of principles training and specific training using several types of training devices. SPECDEVCECN 269-7-102, Contract N6onr-269, Pennsylvania State University, 1955, AD91944.


Weitz, J., & Adler, S. The optimal use of simulation. New York University, Department of Psychology, 1971, AD730951.


Wickens, D. D., & Cotterman, T. E. The effect of display interruption on transfer of training between tasks on different control sensitivity. Ohio State University, Columbus, Ohio, 1958, AD131027.

Wiener, E. L. Transfer of training from one monitoring task to another. Ergonomics, 1967, 10(6), 649-658.


Williams, A. C., Jr. Preliminary analysis of information required by pilots for instrument flight. SDC71-16-1, Contract N60r1-71, University of Illinois, 1949, AT170739.


