DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.
ABSTRACT

An exploratory study was undertaken, as part of a program to develop quantitative techniques for prescribing the design and use of training systems. As a first step in this program, the present study attempted to: (1) compile an initial set of quantitative indices, (2) determine whether these indices could be used to describe a sample of trainee tasks and differentiate among them, (3) develop a predictive methodology based upon the indices, and (4) assess that methodology.

The compilation included the Display-Evaluative Index, a set of panel lay-out indices, and a set of task rating scales. These indices were applied to task analytic data, collected on sonar operator trainers at Fleet Sonar School, Key West, Florida. Application of the indices proved feasible, and differentiation among three training devices, and within four trainee sub-tasks (set-up, detection, localization, classification) was possible.

The predictive method which was generated was an adaptation of the standard multiple regression model. Mean task scores replaced the usual individual criterion scores, and quantitative task index values were used as predictor scores. This adaptation was tested using data from published studies on tracking. Significant multiple correlations using task indices were found for criterion data obtained during early stages of practice. This result supported the contention that a prescriptive method must include training as well as task indices in order to account for advanced levels of proficiency. A combination of task and training indices did predict later performance.

More generally, the results support the soundness of the task characteristic approach underlying the broader program. A major conclusion was that further development of the quantitative task-analytic methodology is warranted and would be fruitful.
INTRODUCTION

The prediction of human performance is highly task specific. Human information processing requirements, stress and overloadings, vary from task to task. Although some of the characteristics of the tasks that create such differences are known, quantitative indices of task characteristics for use in predicting those differences in human performance are generally lacking, especially with respect to the joint influence of several task characteristics as found in complex military tasks. Without quantitative information relating human performance to task characteristics, such things as the instructor's performance level and the trainee's learning rate in a new trainer are difficult to estimate until that trainer is operative.

PURPOSE

The objective of this research project is to develop qualitative indices of the characteristics of instructors' and trainees' tasks so that the effectiveness of a given amount and type of training on a given task can be predicted. The results of this research should lead to greater accuracy in establishing the human performance requirements in a training system, greater accuracy in human factors design recommendations, and improved instructor station design.

In the first phase of this research project—described in this technical report—the objective was to develop a method for quantifying the tasks involved in training device situations, utilizing indices and techniques previously developed and reported in the literature.

ACCOMPLISHMENTS

The objective of the first phase was accomplished very successfully. The initial set of quantitative indices was tested for its feasibility in describing the trainee tasks on three naval operator trainers and in predicting trainee performance. Feasibility was demonstrated.

Further, the feasibility of using quantitative task characteristic indices to predict performance was tested by describing the characteristics of tracking tasks appearing in the experimental literature and predicting tracking performance.

The success of the first phase reinforces the decision to continue the attempt to develop quantitative profiles of training device tasks and to predict performance from such profiles.

FUTURE

The planned next step is to apply quantitative indices to tasks and instructor tasks of actual training device situations and to predict performance in those situations.
It is believed that the development and validation of this type of methodology will make it possible to answer such questions as the following:

(a) What is the relative difficulty of operation of alternate equipment systems?

(b) How long will it take the instructor or trainee to learn the task?

(c) How well is the task carried out by the instructor or trainee?

(d) What is the optimum trainee-to-instructor ratio?

(e) How is the effectiveness of a given amount and type of training on a given task?

[Signature]

GENE R. BICKERT, Ph.D.
Naval Factors Laboratory
ACKNOWLEDGMENTS

The focus of this project was on the "real world", a setting in which the best laid research plans are often modulated by reality. The ability of this project to meet and overcome the problems associated with research "in the field" was due in large measure to the support of the Project Monitor, Dr. Gen. S. Michel. The project staff wishes, therefore, to thank Dr. Michel for his cooperation and able assistance, factors which contributed directly to the success of this project.

Research conducted during the course of "site visits" can often be a frustrating affair. Site-visits made during the present study, however, were highly rewarding. This was primarily due to the interest and unusually able assistance provided by the officers and enlisted personnel of the U. S. Navy Fleet Sonar School, Key West, Florida. Individuals too numerous to mention gave unstintingly of their time and energy. The project staff would like to extend special thanks to Capt. Sanders, Commanding Officer; Cmdr. Elborn, Executive Officer; Lt. Cmdr. Jones; Lt. "Tennant; Chief Petty Officer Morton; and Petty Officers Smith, Stain, and McRae.

The authors, of course, did not do all of the work reflected by this report. We would like to take this opportunity, therefore, to acknowledge the contributions of other project members. Thanks are extended to Dr. William J. Baker of NR who developed much of the rationale underlying the predictive model used in this study. Thanks are also extended to Miss Susan Emery of NP. Susan made valuable contributions throughout the course of the project. Last, but by no means least, thanks are extended to Mrs. Lily Gainer, project secretary, for her patience and able assistance.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Methods for Increasing Training Effectiveness</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Research Related to Training Effectiveness</td>
<td>5</td>
</tr>
<tr>
<td>1.3 Statement of the Problem</td>
<td>8</td>
</tr>
<tr>
<td>2.0 METHODS</td>
<td>10</td>
</tr>
<tr>
<td>2.1 TASK ANALYSIS</td>
<td>10</td>
</tr>
<tr>
<td>2.1.1 Selection of Training Devices</td>
<td>10</td>
</tr>
<tr>
<td>2.1.2 Identification of Trainer Sub-Tasks</td>
<td>12</td>
</tr>
<tr>
<td>2.1.3 Selection of Quantitative Indices</td>
<td>13</td>
</tr>
<tr>
<td>2.1.3.1 Rating Scales</td>
<td>14</td>
</tr>
<tr>
<td>2.1.3.2 Display Evaluative Index</td>
<td>15</td>
</tr>
<tr>
<td>2.1.3.3 Panel Lay-Out and Task-Type Indices</td>
<td>16</td>
</tr>
<tr>
<td>2.1.3.4 Miscellaneous Generic Indices</td>
<td>16</td>
</tr>
<tr>
<td>2.1.3.5 Specific Indices</td>
<td>17</td>
</tr>
<tr>
<td>2.1.4 Application of Indices</td>
<td>18</td>
</tr>
<tr>
<td>2.1.4.1 Display Evaluative Index (DEI)</td>
<td>23</td>
</tr>
<tr>
<td>2.1.4.2 Panel Lay-Out and Task-Type Indices</td>
<td>38</td>
</tr>
<tr>
<td>2.1.4.3 Additional Indices</td>
<td>36</td>
</tr>
<tr>
<td>2.2 Predictive Methodology</td>
<td>56</td>
</tr>
<tr>
<td>2.2.1 Regression Model</td>
<td>57</td>
</tr>
<tr>
<td>2.2.2 Selection of Tasks</td>
<td>59</td>
</tr>
<tr>
<td>2.2.3 Criterion Measures</td>
<td>41</td>
</tr>
<tr>
<td>2.2.4 Application of Indices</td>
<td>42</td>
</tr>
<tr>
<td>2.2.5 Data Analysis</td>
<td>45</td>
</tr>
<tr>
<td>3.0 RESULTS AND DISCUSSION</td>
<td>44</td>
</tr>
<tr>
<td>3.1 Task Analysis</td>
<td>44</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS (Continued)

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.1 Generic Indices</td>
<td>41</td>
</tr>
<tr>
<td>3.1.1.1 Display Evaluative Index</td>
<td>41</td>
</tr>
<tr>
<td>3.1.1.2 Panel Lay-Out and Task-Type Indices</td>
<td>47</td>
</tr>
<tr>
<td>3.1.1.3 Miscellaneous Generic Indices</td>
<td>49</td>
</tr>
<tr>
<td>3.1.1.4 Task Characteristic Rating Scale</td>
<td>50</td>
</tr>
<tr>
<td>3.1.2 Specific Indices</td>
<td>50</td>
</tr>
<tr>
<td>3.2 Prediction Studies</td>
<td>52</td>
</tr>
<tr>
<td>3.2.1 Rating Scale</td>
<td>52</td>
</tr>
<tr>
<td>3.2.2 &quot;Human Engineering&quot; Indices</td>
<td>55</td>
</tr>
<tr>
<td>3.2.3 Combined Indices</td>
<td>56</td>
</tr>
<tr>
<td>3.2.4 Discussion</td>
<td>58</td>
</tr>
<tr>
<td>4.0 CONCLUSIONS</td>
<td>62</td>
</tr>
<tr>
<td>5.0 RECOMMENDATIONS</td>
<td>64</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>66</td>
</tr>
<tr>
<td>APPENDIX A Task Characteristic Rating Scales</td>
<td>70</td>
</tr>
<tr>
<td>APPENDIX B Operations Flow-Charts</td>
<td>85</td>
</tr>
<tr>
<td>APPENDIX C DII Link Charts</td>
<td>99</td>
</tr>
<tr>
<td>APPENDIX D Prediction Study References</td>
<td>113</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Activity Table for 14E10 Localization Sub-Task</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>Equipment Function Table for 14E10 Location</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>Link Table for the 14E10 Localization Sub-Task</td>
<td>28</td>
</tr>
<tr>
<td>4</td>
<td>Link Worksheet for the 14E10 Localization</td>
<td>29</td>
</tr>
<tr>
<td>5</td>
<td>Link Value Table for the 14E10 Localization</td>
<td>31</td>
</tr>
<tr>
<td>6</td>
<td>Generic &quot;Human-Engineering&quot; Indices Derived From Task Analysis Data</td>
<td>45</td>
</tr>
<tr>
<td>7</td>
<td>Task Characteristic Ratings: Derived From Task Analysis Data</td>
<td>51</td>
</tr>
<tr>
<td>8</td>
<td>Multiple Regression Analyses</td>
<td>53</td>
</tr>
</tbody>
</table>

LIST OF ILLUSTRATIONS

<table>
<thead>
<tr>
<th>Figure</th>
<th>14E10 Localization Sub-Task Operations Flow Chart</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>14E10 Localization Sub-Task Link Chart</td>
<td>26</td>
</tr>
<tr>
<td>3</td>
<td>14E10 Localization Sub-Task Panel Lay-Out Diagram</td>
<td>33</td>
</tr>
</tbody>
</table>
One of the most difficult and complex problems faced by those individuals responsible for military training is the development of effective training systems. Accepted as inputs to such systems are students who lack specific knowledge or skills. As the students progress through the system, they are exposed to a variety of new situations, and through repeated use of similar opportunities to develop and refine new skills. In military training systems, the bulk of this exposure is often provided by sophisticated training devices which incorporate conditions intended to facilitate, guide, and reinforce the learning experience. The objective of supplying such exposure is to output graduates who possess new knowledge and skills, and who can transfer these assets to an operational situation. The effectiveness of this process, and of the system components which underlie it, depends largely upon the speed of skill acquisition, and the degree to which these skills transfer to the operational setting.

While the goal in designing military training devices is always one of maximizing effectiveness, any given device may fall short of this mark. Two devices having the same training objectives can differ markedly in terms of the training time required or the type and amount of transfer achieved. Here a choice available between prototypes of these devices, only one might be selected for development. Because of the costs involved in current system design and implementation, however, one cannot afford to develop competitive devices. and empirically determine which is more effective. In other words, a "hunch-guess" attitude about the effectiveness of training is imperative.

The basic issue, therefore, is how to plan, design, and develop a training device which will prove to be effective for a particular set of training objectives. Given the complexity and
training, how can one forecast the type of training device which should be employed? How can one specify the purpose in which the device should be designed and utilized? In short, what steps can be taken to insure, insofar as possible, the rapid acquisition of skills and their positive transfer to the operational setting?

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

1.1 Methods for Increasing Training Effectiveness

Historically, gross inadequacies in the design of training devices were often eliminated on the basis of shrewd guesswork. In the earliest approach to the prescriptive-predictive issue, design decisions were made by subject-matter specialists who drew on experience and common sense to solve training design problems. As a result they often were able to make fairly sound decisions about the design of training aids and equipment, student and instructor stations, and other aspects of the training situation which might facilitate the learning experience. However, these early practitioners were artisans. Because of their experience, they were able to translate certain types of information about the job to be
performed into requirements for training. As in the majority, however, they differed in terms of their conceptualization of an effective approach to the training problems which faced them. As a result, they were highly successful in making sound training decisions, others were not. Furthermore, because of the informal and intuitive nature of their methods, it was difficult to test others in terms of what the principal disadvantage of this approach lay in the difficulty of evaluating the proposed training solution prior to its adoption. Predictions as to the effectiveness of training were scarcely better than opinion. As has been cogently pointed out elsewhere (Cheroff & Miller, 1965):

"One can never know, about any training program so devised, whether all of the important aspects of the man-machine system have been considered, whether training has been prescribed for those aspects which bear the greatest relationship to system effectiveness, and whether the training program is particularly well suited to teaching the specific skills and knowledge which must be acquired. It is extremely difficult to assess, before the fact, whether each training dollar will be well spent" (p. 2).

Because of the difficulties inherent in these individualistic methods, attention was focused upon the development of more formal and programmatic approaches to the solution of training device design and utilization problems. Techniques were conceived and developed to help determine what should be trained and to provide very general guidelines as to how training should be accomplished.

The results of these efforts were a number of job descriptive and task analytic procedures. Using these approaches it became possible to describe jobs in terms of their major task components, and then to describe these components in terms of underlying task elements and activities. Description proceeded systematically through several levels.
At each successive level the information which was extracted became more detailed. Additional techniques were employed to either expand upon or integrate most of this information. The earliest of these procedures (e.g., Miller, 1963; Miller & Van Tutt, 1955) were designed to help specify those aspects of an operational task which should be considered as basic items of content in a training program. More recent efforts (e.g., Chownell & Volle, 1965), while maintaining an interest in specifying the approximate content of training, have also attempted to preclude the time and amount of training which should be given.

Concurrent with these activities, other investigators were attempting to develop task classification systems having implications for training. These taxonomists shared the belief that basically different types of tasks did indeed exist. Given this premise, a logical step was to collect, sort, and catalogue tasks, casting them into their appropriate classes or families. Taxonomists who developed their structures to deal with training problems made an additional assumption. For each identifiable class of tasks there might exist an unique set of training procedures which could prove to be most effective. As a consequence of this thinking there have been several attempts to classify tasks and to specify for each class those training techniques which seemed most appropriate (e.g., Wills & Peterson, 1961; Stolswok, 1964; Miller, 1966).

Many of the analytic and taxonomic methodologies developed to date have had their own particular shortcomings. Most, however, have had one weakness in common. They have provided for the description of tasks in behavioral or functional terms (e.g., the behavioral taxonomy of Berliner, Angell, & Shearer, 1961; the functional descriptors employed by Gagne, 1962 and Miller, 1966). These terms have been found difficult to apply unambiguously. They have referred primarily to the qualitative aspects of an operator's overt and covert behaviors.
Considered collectively, these qualitative approaches have helped determine those aspects of the operational situation which should be considered as basic items of content in the training program. To a much more limited extent, they have even provided general guidelines about how training might best be accomplished. They have not been particularly successful, however, in establishing specific training device design requirements. As pointed out by Smith (1965) in an interesting comparative study, it has been difficult to translate the behavioral analyses into rigorous training technique or hardware specifications. Because of their qualitative nature, it has not been possible to use these methods to predict the effectiveness of different types or amounts of training.

1.2 Research Related to Training Effectiveness

The problem of maximizing training device effectiveness is twofold: 1) to predict, as early as possible in the design process, how effective training will be; or 2) to specify that design, which if carried through, will prove effective. In either case, the problem is overwhelmingly complex, since in any training situation there are several major classes of variables which may interact to determine the rate of trainee skill acquisition or even the level of instructor proficiency. These components are the trainees who are selected, the characteristics of the tasks embodied in the training device, and the techniques employed to effect training.

In mentioning these particular components an important point is to be made. If methods for predicting the effectiveness of a given type and amount of training are to be developed, the complex interactions among these components must be investigated. Studies of this type, however, have been relatively scarce. While the prediction of learning rate or performance level has intrigued behavioral scientists
for a number of years, the tendency has been to focus on separate aspects of the problem. These divergent interests have been reflected in the growth of three rather separate areas of research. These areas have included individual differences, principles of learning, and human engineering.

The first of these areas has been aimed at determining student-related attributes which may underlie individual differences in training. The value of this research lies in the promise it holds for using such student attributes as a basis for student selection. The rationale for this selection is that the most effective training can be provided to those students who possess attributes related to rapid acquisition of certain skills or to high levels of proficiency on certain types of tasks.

A number of studies, for example, have indicated that the abilities derived through experimental-correlational research are involved in varying degrees and combinations in learning to perform a variety of tasks. Among others, Fleishman and Fruhtr (1960) and Parker and Fleishman (1960, 1961) have applied this knowledge while providing training on a variety of complex tasks. Few of these studies, however, have attempted to map the relationships between different student attributes and different tasks or the stimulus and response properties of such tasks. Until this research is undertaken on a large scale, it will be difficult to predict which students will gain most from practice on different types of tasks. Lacking this information it is difficult to determine how much device effectiveness can be increased through personnel selection.

The second relevant area has focused on the development of principles of learning and the translation of these principles into sound training practices. These have included, for example, the work of Bilodeau and Bilodeau (1961) on feedback, the distribution of practice studies by Eientl (1961) and Jensen (1961), Cofer and Appley's (1964) work on motivation, and Caro's analysis (1969) of adaptive training.
In spite of the investigation of a large number of such variables, however, it has been difficult to translate this body of research into the design of improved training devices.

This difficulty has arisen for three reasons. First, few experiments have been conducted to explore the interactions among different training variables. In the absence of this type of research it has been difficult to evaluate device effectiveness in terms of trade-offs among training variables. Second, there have been few attempts to establish the interactions between selected training variables and student attributes or abilities. Third, with the exception of a few studies (e.g., Tallmadge and Shearer, 1970) the interactions between selected training variables and the types of tasks being trained have not been thoroughly explored. More of this latter type of research is needed if training techniques are to be tailored to the tasks incorporated within a particular training device.

The third relevant area of research has focused on stimulus and response aspects of a task, which if varied, may exert an effect upon operator performance. The research in this area has been voluminous. It has led to the generation of a number of handbooks prescribing the design of most aspects of the man-machine interface (e.g., the Human Engineering Guide to Equipment Design, 1965; the Handbook of Human Engineering Data, 1960). While of immense value in its present form, this research must be extended. More studies of the type conducted by Chapanis and Gropper (1968) are required on the interaction between operator characteristics and display-control relationships. Similarly, more information is needed of the type supplied by Fowler, Killian, Fowler, and Young (1968) on different operator panel lay-outs and rates of learning.

In order to cope successfully with the prescriptive-predictive issue, human-engineering research is especially needed on the interaction between trainee and instructor. At his station, the trainee attends to
inputs from specific displays, processes them in a prescribed manner, and attempts to take appropriate control actions. The instructor, at his own station, deals with different controls and displays, and performs distinctly different operations. To some extent, however, each station's output is a function of inputs from the other station. Because of this dependency, a training device is actually a closed-loop system. Therefore, device effectiveness will depend upon how well both participants perform their respective tasks. Consequently, design prescriptions must simultaneously relate to both stations, and predictions of training effectiveness must consider their joint influence.

In summary, the three areas of research described above indicate that training effectiveness is determined by an interaction among components of the training system. Study of this interaction may eventually provide the design engineer with information about the personnel, technique, and design trade-offs which are so crucial to development of an effective training device. Until the relationships among these components are thoroughly understood, the problem of designing an effective training device, before the fact, will remain unresolved.

1.5 Statement of the Problem

The present study was conducted as part of a larger program of research. The goal of this program was development of a new approach to the problem of specifying the design of a training device or of predicting its effectiveness. The approach under consideration was one of quantitative task analysis.

As the first step in this program the present study had three objectives. The first objective was to compile an initial set of quantitative indices relating to selected characteristics of various man-machine tasks. The second objective was to determine whether the obtained indices could be used to describe a sample of trainee tasks and
to differentiate among them. The third objective was to develop a predictive methodology based upon the task indices and to assess its potential utility.


2.0 METHOD

Two methods were employed in the present study. In the first procedure quantitative task indices were compiled and applied to trainee tasks found in surveillance-system training devices. The purpose of this procedure was to assess the feasibility of quantitatively describing a variety of complex non-machine tasks. In the second procedure a multiple-regression model was developed and applied to a sample of tracking tasks described in the literature. This procedure was designed to provide preliminary estimates of the predictive power of selected task indices. The major steps in both approaches are described below.

2.1 Task Analysis

Four major steps comprised the task analysis procedure. First, a sample of training devices was selected upon which to base the eventual quantitative task analysis. Second, the trainee tasks associated with the selected devices were analyzed to identify major sub-tasks believed to cut across a number of training devices. Third, quantitative indices were selected or developed relating to characteristics of the major trainee sub-tasks which had been selected for study. Fourth, task analysis data were collected in the field and used to derive values for both generic and specific sets of quantitative indices.

2.1.1 Selection of Training Devices

A large portion of the spectrum of Navy training devices was reviewed in order to identify those instances in which training equipments rather than training aids provided the basis for instruction. The former devices (e.g., trainers and simulators) were chosen for investigation because: 1) they contained trainee and instructor tasks which were reasonably formalized and invariant with respect to the equipment.
and procedures used; and 2) they permitted relatively sharp boundaries to be drawn between the trainee and instructor tasks. Both of these features were desirable for development and application of quantitative indices.

On the basis of the review approximately 165 different trainers or simulators were identified. These equipments differed markedly, however, in terms of the basic content of training (e.g., vehicle control, fire-control, navigation, etc.) and level of training (e.g., orientation, familiarization, skill, etc.). A decision was required, therefore, whether to sample across these many different types of trainers or to focus on a more homogeneous sub-set of devices. The latter approach was finally adopted because it was felt that focus on a specific sub-set of devices would provide a better test of the overall methodology. If quantitative indices could not be applied to a specific class of trainers, then there would be little hope of doing so across many different types of devices.

The 165 devices previously identified were re-evaluated with respect to the content of training, and were organized into relatively homogeneous families. Nine clusters emerged which included the following:

a) operational flight trainers  
b) cockpit procedures trainers  
c) weapon system trainers  
d) antisubmarine warfare team trainers  
e) airborne electronic warfare trainers  
f) electronic countermeasures trainers  
g) radar trainers  
h) sonar trainers  
i) miscellaneous trainers.

Two sets of criteria were applied to each group of devices. One set was designed to identify groups of devices which appeared best suited to
development of quantitative indices. The second set was used to specify which groups might be most readily evaluated in an anticipated regression analysis.

In light of the criteria for development of indices and for evaluation of the descriptive system, radar, sonar, and electronic countermeasures trainers were selected. These devices belonged to the same general family in the sense that they provided training for the operators of active or sensor-based surveillance systems. For purposes of the present study, attention was focused on active, surface-sonar trainers. In spite of this restriction, the intention was to generate indices which would also provide for the quantitative description of other devices within the surveillance family.

2.1.2 Identification of Trainee Sub-Tasks

Having identified surveillance system training devices as the family of interest, the trainee tasks associated with these devices were analyzed in detail. The analysis was conducted for two reasons. First, information was required on the major sub-tasks performed by trainees. Second, information was desired about those features of the sub-tasks which might provide a basis for generation of descriptive indices.

The decision to provide description at the sub-task level was predicated on two assumptions. First, although surveillance trainers might differ in the content of training, they nevertheless would share certain basic sub-tasks. Second, only at the sub-task level could criterion performance measures be readily identified. The availability of such measures was essential if the quantitative indices were eventually to be used in the prediction of learning rates or proficiency levels.
Ten devices were evaluated during site visits to the Fleet Sonar School, Key West, Florida, and the Naval Air Station, Glynn, Georgia. Based upon this analysis and upon examination of a number of utilization manuals, four, major, trainee sub-tasks were identified which cut across surveillance training devices. The first sub-task was procedural in nature and involved receiver turn-on, set-up, and/or calibration in preparation for search activities. The second sub-task, involving monitoring of the receiver, resulted in signal detection or target acquisition. In the third sub-task, displayed signals were analyzed to permit target identification and classification. The fourth sub-task involved tracking of the target in order to provide continuous or discrete information about target range and bearing. All four sub-tasks were readily identifiable in the active, surface-sonar devices with which the study was primarily concerned.

### 2.1.3 Selection of Quantitative Indices

In selecting and developing a set of quantitative indices there was an embarrassment of riches. Once compilation of the list of descriptors began it was all but impossible to stop. To combat this excess a line had to be drawn somewhere. Consequently, quantitative indices were sought which related only to critical characteristics of each of the trainee sub-tasks identified above. Critical characteristics were those which, if manipulated, could be hypothesized to exert an appreciable effect upon rate of acquisition or level of proficiency.

Based upon a review of the literature and upon an examination of the four trainee sub-tasks of primary interest, two sets of indices were generated. The first set consisted of generic indices. Each index within this first set was applicable to all of the trainee sub-tasks as well as to the task of the instructor. The generic indices included: 1) task characteristic rating scales; 2) a display evaluative index; and 3) a variety of panel lay-out and task-type indices. The
second, a second set of specific indices which were developed to provide a more detailed description of each of the trainee sub-tasks. An index within this second set was specific in the sense that it would apply to at least one, but not to all, of the trainee sub-tasks.

2.1.3.1 Rating Scales: A total of 13 task characteristic rating scales was selected from a larger set of 19 scales originally developed during the course of an MP language project (Melnick, Teichner, and Stephenson, 1970). The scales were specifically designed to describe tasks per se, independent of the other major components of performance, the operator and the task environment. Development of the scales proceeded from a definition which structured the term "task" into several components: the goal, resources, procedures, stimuli and stimulus-response relationships. Several rating scales were developed for each of these components, and small-scale studies were performed to assess inter-judge reliabilities. A complete discussion of the task characteristic approach is given in a report by Laring and Wheaton (in press).

For purposes of the present study the rating scales were reassessed with the surveillance sub-tasks in mind. Where possible, a change was made from rating the magnitude of a characteristic (on a seven-point scale) to actually counting the quantity involved. For example, rather than rating the number of responses required to produce an output unit, such responses were counted. Consequently, the resulting instrument was a mixture of rated and counted characteristics.

Although all of the scales provided quantitative information, a few of them were actually based on qualitative distinctions. In these cases the different qualitative states were assigned arbitrary values on a nominal scale. Definitions for each task characteristic and the associated rating scales are presented in Appendix A. The task characteristic indices employed in the present study included the following:
a) Number of output units (OPU)
b) Duration (DURA)
c) Number of elements per output unit (ELPU)
d) Work load (LOAD)
e) Precision of responses (PRPC)
f) Simultaneity of responses (SIMU)
g) Number of responses (NOR)
h) Rapidness of feedback (FRF)
i) Response rate (RRAT)
j) Tutorial dependency (TIDE)
k) Natural dependency (NIDE)
l) Operator control of the response (OCR)
m) Variability of stimulus location (VARS)

2.1.5.2 Display Evaluative Index - Over the last decade a number of attempts have been made to translate human engineering principles into quantitative measures. Among these has been Siegel's Display Evaluative Index (DEI).

The DEI is a measure of the effectiveness with which information flows from displays via the operator to corresponding controls. The index yields a dimensionless number which represents a figure of merit for the total configuration of displays and controls being evaluated. It was originally derived from a set of assumptions about what constitutes efficient information transfer in display-control systems. For example, all else being equal, that system is best which has the greatest directness between the information transmitter and the corresponding controls, efficiency being reduced where the operator has to transform information before taking action.

The potential value of the index has been demonstrated by its wide applicability. Surveillance, fire-control, and even communications systems have been quantified with it (e.g., Siegel, Hichle, Stedronian.
Moreover, the index has been partially validated, i.e., against judgments by human engineering experts (Siegel, Michie, & Federman, 1963, 1968). It was decided, therefore, to include the PDI in the current project with a view toward extending its prescriptive-predictive potential to the design of training devices.

2.1.4. Panel Lay-out and Task-Type Indices - The indices of Tolley, et al., (1968) are designed to provide description of two different aspects of a man-machine task, one set of indices is used to measure, in percentage, the extent to which general human engineering principles have been applied to the arrangement of controls and displays on a console. The second set relates to the degree, again expressed as a percentage, to which different operations or "task types" are embodied in a particular operator console. These indices can vary independently of the PDI which does not address itself to panel arrangements or types of panel operations.

In the present study one lay-out index was used, the "total sequencing score". Four "task-type" indices were employed including the following: 1) an "alternative action sub-score"; 2) a "breaks in operation sequence sub-score"; 3) a "frequency-of-use score"; and 4) an "importance-of-use score". In addition to these major indices, certain measures involved in their calculation were also used as descriptors. These included: 11 total link value; 21 number of controls and displays; and 31 total number of response actions.

2.1.4.1 Miscellaneous Generic Indices - To round out the initial set of generic indices, seven additional measures were employed. Response actions were broken down into the following categories: 1) number of non-normal repertoire responses (Tolley, 1964); 2) number of control activation responses; 3) number of feedback responses; 4) number of information acquisition responses; and 5) number of instructor initialized responses.
(Mackie & Narabedian, 1967). Two additional indices were the number of 
redundant information sources processed simultaneously (Mishell, 1965), 
and the time permitted for sub-task completion.

With the inclusion of the seven indices just described, the generic 
set consisted of 29 separate measures. This set was deemed acceptable for 
initial work in terms of both the number and variety of descriptors which 
were available.

2.1.3.5 Specific Indices - In addition to generic indices, which 
cut across both training devices and trainee sub-tasks, an additional 
set was selected. Indices within this set were specific to surveillance 
trainers and to certain sub-tasks within those trainers. The items were 
selected because they appeared to have implications for device design 
decisions and because they appeared to be directly translatable into 
trainer design specifications.

The 15 specific indices developed for use in the present study in- 
cluded the following:

1) signal persistency expressed as the ratio of phosphor 
persistence to the ping interval;
2) range in signal to noise ratios;
3) bearing control-display ratio;
4) range control-display ratio;
5) number of tracking dimensions;
6) variation in target range;
7) variation in target speed;
8) bearing error tolerance;
9) range error tolerance;
10) number of cues available for classification;
11) number of classification cues applied simultaneously;
12) number of false targets used;
2.1.1 Application of Indices

The general and specific indices discussed above were applied to task-analytic data collected on three sonar operator training devices in use at the Fleet Sonar School, Key West Naval Base, Florida. The three trainers examined were: the USNS-10, representing the W/SOS-15 helicopter stack, the W/SOS-20X, and the USNS-1, representing the W/SOS-1 sonar. The "US" and the "S" are destroyer systems. All three systems have at least some capability for detection, localization, and classification of submerged contacts and were regular assignments to the trainers. Included in the operation of each surveillance system is the sequence taught to novice operators. Considerable care was taken both by the instructors and by the observers to maintain a training rather than an operational set.

Procedures for equipment set-up, detection, localization, and classification were included in each demonstration. For each of these sub-tasks the instructor indicated and described every display and control used and their sequence of use. This information was recorded on three forms: 1) an activity table describing the actions performed; 2) an equipment function table describing the displays and controls; and 3) an operational...
sequence diagram. There was some deliberate redundancy among the data forms. These respective forms are illustrated for the IOTD basic action task in Tables 1 and 2, and in Figure 1. Procedures for checking the remaining tasks and devices are presented in Appendix C.

Table 1 shows main line actions on the left and continuance actions on the right. Table 2 describes the displays and controls corresponding to each line of action in Table 1. Table 2 includes response number, equipment reference number, designation of equipment as a control (C), display (D), or a combination of both (CD). Also included are equipment category name and number of hardware units (number of discrete values which can be read out of a display or entered into a control). The "In response repertoire" column indicates whether the specific action required is part of the trainee's normal repertoire or whether it represents a skill to be acquired. The "Feedback" column represents responses in which the adequacy of a preceding response is confirmed. The "Importance" column indicates the criticality of an erroneous response. A "1" rating indicates that the training mission would be inhibited but the error is correctable. A "2" rating indicates that the mission would fail but the error is correctable. A "5" rating indicates maximum criticality, i.e., a wrong response results in damage to the equipment. The "Control by instructor" column indicates instances in which the instructor manipulates a control for the trainee or tells the trainee to enter a specific value into a control.

Figure 1 represents an integration of information contained in Tables 1 and 2. The sequence of actions required in the localization sub-task is shown graphically from left to right. Squares denote actions involving use of a display while circles denote actions involving controls. The main sequence of responses is represented by a solid line between controls and displays while continuance response sequences are indicated by broken lines. Controls and displays are arbitrarily oriented, starting from the left, to indicate when they are used in the sequence of responses.
### Table 1
Activity Table for 14010 Localization Sub-Task

<table>
<thead>
<tr>
<th>Resp. #</th>
<th>Response Description</th>
<th>Contingency Response #</th>
<th>Contingency Response Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Monitors target on display #22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Rotates control #10 and #23 to align cursor with target</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Monitors position of cursor relative to target on display #22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Reads ranging data from display #20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Reads range data from display #20a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Depresses foot pedal (#30) to call pilot</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Receives go ahead from pilot over earphones (#20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Uses microphone (#31) to report range and bearing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Receives hearing &quot;clear&quot; or &quot;foul&quot; information from #20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.1.a</td>
<td>If hearing is &quot;foul&quot; the search sub-task is reinitiated by rotating control #6,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Monitors #22 for target motion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Rotates control #10 and #23 to keep cursor on target</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Monitors position of cursor relative to target on display #22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Checks #20a for range relative to previous range setting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.1.a</td>
<td>If range is increasing or decreasing, manipulates #11 to select a new range scale</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resp. #</td>
<td>Response Description</td>
<td>Contingency Response</td>
<td>Contingency Response</td>
</tr>
<tr>
<td>--------</td>
<td>--------------------------------------</td>
<td>----------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>11</td>
<td>Monitors #20 for target</td>
<td>13.1.b</td>
<td>13.1.c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13.1.c</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>14.1a</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>14.1b</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>14.1c</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Monitors #22 for target</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Rotates #19 and #23 to track target</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Monitors #22 for target</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resp.</td>
<td>Equip. #</td>
<td>Equip. Nomencature</td>
<td>Hardware Y Units</td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>-------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>1</td>
<td>22</td>
<td>PPI</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>20,23</td>
<td>Bearing and Range Wheels</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>22</td>
<td>PPI</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>Range Indicator</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>20a</td>
<td>Range Indicator</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
<td>Foot-Pedal</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>29</td>
<td>Earphones</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>31</td>
<td>Microphone</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>29</td>
<td>Earphones</td>
<td>-</td>
</tr>
<tr>
<td>9.1.a</td>
<td>6</td>
<td>Listen Sector Switch</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>22</td>
<td>PPI</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>20,23</td>
<td>Bearing and Range Wheels</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>22</td>
<td>PPI</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>20a</td>
<td>Range Indicator</td>
<td>-</td>
</tr>
<tr>
<td>13.1.a</td>
<td>11</td>
<td>Range K Yards Select Switch</td>
<td>5</td>
</tr>
<tr>
<td>13.1.b</td>
<td>22</td>
<td>PPI</td>
<td>-</td>
</tr>
<tr>
<td>13.1.c</td>
<td>20a</td>
<td>Range Indicator</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>20</td>
<td>Earphones</td>
<td>-</td>
</tr>
<tr>
<td>Resp. #</td>
<td>Equip. #</td>
<td>C,D,B</td>
<td>Equip. Nomenclature</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>-------</td>
<td>---------------------</td>
</tr>
<tr>
<td>14.1.a</td>
<td>22</td>
<td>D</td>
<td>PPI</td>
</tr>
<tr>
<td>111.1.b</td>
<td>8</td>
<td>C</td>
<td>Reception Dist. Switch</td>
</tr>
<tr>
<td>14.1.c</td>
<td>29</td>
<td>D</td>
<td>Earphones</td>
</tr>
<tr>
<td>15</td>
<td>22</td>
<td>D</td>
<td>PPI</td>
</tr>
<tr>
<td>16</td>
<td>10,23</td>
<td>C</td>
<td>Parking &amp; Range Whee</td>
</tr>
<tr>
<td>17</td>
<td>22</td>
<td>D</td>
<td>PPI</td>
</tr>
</tbody>
</table>
Figure 1. PHM: Formulation Unit-Tank Operations Flow Chart
2.1.4.1 Display Evaluation Index (DEI) formula is (Siegel, Michael, & Feuerstein, 1969):

\[
(n + m)_{u} \left\{ \exp\left(\frac{1}{1 + \frac{n}{m}}\right) \right\} \\
(1 + \frac{n}{m})^{\frac{1}{m + n}}
\]

where:
- \( n \) = number of indicators (i.e., displays);
- \( m \) = number of controls;
- \( N \) = the number of forward links, i.e., dotted lines from controls to displays are not counted (Figure 2);
- \( (n + m)_{u} \) = number of indicators and controls actually used on the console during a particular sub-task;
- \( (n + m)_{t} \) = total number of indicators and controls on the console;
- \( \Sigma W \) = sum of weights applied to indicator-control links and to rectangles representing operator data processing; weighting procedures are defined in Siegel, et al. (1969);  
- \( \Sigma/M \) = sum of absolute values of mismatches between indicator and control resolution;
- \( Q \) = total number of display and control elements for and controls and displays (e.g., in the 1410, localization is accomplished by manipulating separate bearing and index charts, which can be thought of as two elements of a single control);
- \( n_{o} \) = number of boxes and triangles representing processes that intervene between the reading of indicators and the population of controls, where triangles are used to represent "and" gates and "or" gates.

Following the procedure outlined in Siegel, et al. (1969), analytic data for each sub-task of each training device were reduced to information transfer charts. From the charts, tables of links were established. Then the appropriate formula sums were obtained and the DEIs calculated. These procedures are illustrated in Figure 3 and in Tables 3 and 4. (Link charts are presented for the remaining tasks and devices in Appendix C.)
Figure 2. LMD Localization Sub-Task Link Chart
Figure 2 shows displays on the left, controls on the right and intervening process symbols in the center. Connecting these symbols are links which indicate the direction of data flow. Solid lines extending from display or processing symbols are instruction links. These imply actions based upon display reading or data processing by the operator. The short solid lines which terminate on a processing or control symbol are instructional links. These imply actions taken on the basis of instructions or stored information. The dotted lines extending from controls to displays are corroborative (i.e., feedback) links. These imply that the operator is checking the effects of some control action.

The number within the display and control symbols indicates the number of different values which can be read or manipulated. For example, 122 is a PPI which, for purposes of adjusting the bearing and range wheels, takes on a total of nine values, i.e., the cursor can be short of the target, beyond the target, leading the target, lagging the target, on the target, leading the target and short, leading the target and long, lagging the target and short, or lagging the target and long. These nine states are associated with the links going to the upper two "compare" boxes. However, for purposes of setting the reception direction control, the PPI takes on three values, i.e., the search sector will either coincide with the target sector, lead it or lag it. These three states are associated with the link going to the "and" gate.

The number of elements per display or control is indicated in brackets next to the control or display symbol. The intervening symbols include three compare boxes and two "and" gates. The upper two rectangles for example, designate that the trainee compares cursor position in bearing and range with target position. The "and" gate designates that the trainee combines video and and a information to set the "reception direction" switch.
<table>
<thead>
<tr>
<th>Link No.</th>
<th>Link Type</th>
<th>Display Info.</th>
<th>Control Info.</th>
<th>Link Mismatch</th>
<th>Link Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No.</td>
<td>States</td>
<td>Digits</td>
<td>No.</td>
</tr>
<tr>
<td>1</td>
<td>INFO</td>
<td>9</td>
<td>.96</td>
<td>9</td>
<td>.96</td>
</tr>
<tr>
<td>2</td>
<td>INST</td>
<td>9</td>
<td>.95</td>
<td>9</td>
<td>.95</td>
</tr>
<tr>
<td>3</td>
<td>INFO</td>
<td>9</td>
<td>.95</td>
<td>9</td>
<td>.95</td>
</tr>
<tr>
<td>4</td>
<td>INST</td>
<td>9</td>
<td>.95</td>
<td>9</td>
<td>.95</td>
</tr>
<tr>
<td>5</td>
<td>INFO</td>
<td>3</td>
<td>.18</td>
<td>3</td>
<td>.18</td>
</tr>
<tr>
<td>6</td>
<td>INFO</td>
<td>5</td>
<td>.48</td>
<td>5</td>
<td>.48</td>
</tr>
<tr>
<td>7</td>
<td>INFO</td>
<td>.961</td>
<td>9</td>
<td>.96</td>
<td>.00</td>
</tr>
<tr>
<td>8</td>
<td>INFO</td>
<td>3</td>
<td>.18</td>
<td>3</td>
<td>.18</td>
</tr>
<tr>
<td>9</td>
<td>INFO</td>
<td>2</td>
<td>.50</td>
<td>2</td>
<td>.50</td>
</tr>
<tr>
<td>10</td>
<td>INFO</td>
<td>2</td>
<td>.50</td>
<td>2</td>
<td>.50</td>
</tr>
<tr>
<td>11</td>
<td>INFO</td>
<td>.782</td>
<td>.48</td>
<td>.48</td>
<td>.50</td>
</tr>
<tr>
<td>12</td>
<td>CORROR</td>
<td>2</td>
<td>.50</td>
<td>2</td>
<td>.50</td>
</tr>
<tr>
<td>13</td>
<td>INFO</td>
<td>72</td>
<td>1.86</td>
<td>72</td>
<td>1.86</td>
</tr>
<tr>
<td>14</td>
<td>INFO</td>
<td>2000</td>
<td>5.48</td>
<td>2000</td>
<td>5.48</td>
</tr>
<tr>
<td>15</td>
<td>INFO</td>
<td>2000</td>
<td>5.48</td>
<td>2000</td>
<td>5.48</td>
</tr>
<tr>
<td>16</td>
<td>INST</td>
<td>2000</td>
<td>5.48</td>
<td>2000</td>
<td>5.48</td>
</tr>
<tr>
<td>17</td>
<td>INFO</td>
<td>3</td>
<td>.48</td>
<td>5</td>
<td>.70</td>
</tr>
<tr>
<td>18</td>
<td>INST</td>
<td>2</td>
<td>.50</td>
<td>2</td>
<td>.50</td>
</tr>
</tbody>
</table>

\[ \Sigma |W_i| = 5.52 \]
\[ \Sigma W_i = 51.5 \] (Each box receives a link weight of 4)

1. Number of display digits for link 7 = number of display digits for link 5 plus link 6.
2. Number of display digits for link 11 = number of display digits for link 8 plus link 10.
Table 4
DEI Worksheet for the 1H/10
Localization Sub-Task

1. \((1 + \sum \mathbf{M}_i)\) = 35.5
2. \((n + m)^u\) = 10.0
3. \((n + m)^t\) = 32.0
4. \(N\) = 17.0
5. \(Q\) = 11.0
6. \(n_0\) = 5.0
7. \((Q + n_0)\) = 16.0
8. \(\sqrt{N(n + m)^t (Q + n_0)}\) = 8.704,00
9. \(\sqrt{N(n + m)^t (Q + n_0)}\) = 95.50
10. Sum \(\mathbf{M}_i\) = .52
11. 1/4 Sum \(\mathbf{M}_j\) = .13
12. \(\exp(-1/4 \text{Sum})\) = .88

\[
\text{DEI} = \frac{(n + m)^u \left[\exp\left(-1/4 \text{Sum} \mathbf{M}_i\right)\right]}{(1 + \sum \mathbf{M}_i) \left[\sqrt{N(n + m)^t (Q + n_0)}\right]}
\]

\[
= \left(\frac{10}{35.5}\right) \left(\frac{.88}{95.50}\right) = \frac{8.8}{3.31} = .0026568
\]
Table 3 describes in detail the links shown in Figure 2. The description includes link type, number of values which can be read from the associated display and/or entered into the associated control, mismatch between these numbers in digits, i.e., absolute difference between number of display digits and control digits, and link weight. The approximate 'N1 sums are indicated at the bottom of the page. Table 1 is a worksheet derived from Siegel, Hickle, and Federman (1962).  

### 2.1.4.1 Panel Lay-out and Task-Type Indices - The task analysis data shown in Table 1 and 2 and in Figure 1 were used to obtain values for eight panel lay-out and task-type indices. Only general methods for deriving index values have been described in the present report. A thorough and detailed description of these procedures has been provided elsewhere (Fowler, et al., 1968).  

Many of the indices developed by Fowler, et al. (1968) are based upon the concept of a link. A link is defined as the hand movement between two controls and the eye movement between two displays or between a display and a control. In Figure 1 links are shown between the displays and controls employed in the 14010 localization sub-task. Links involved in the main sequence of actions are represented by solid lines. Those occurring in contingency sequences are represented by broken lines.  

The first step in deriving many of the indices is to convert the information shown in Figure 1 into a Link Value Table (Table 5). Each link in Figure 1 is listed in coded form in column 1 of Table 5. The first number in the code refers to the display or control from which a given link leaves. The second number refers to the hardware component which the link then enters. In columns 2, 3, and 4 of Table 5, the following data are recorded for each link: 1) the number of times the link is used; 2) the relative percentage of use of a link leaving a given control or display; and 3) a link value which is the product of data
Table 5
Link Value Table for the 14E10 Localization Sub-Task

<table>
<thead>
<tr>
<th>Links</th>
<th>No. Times Link Used</th>
<th>% Use</th>
<th>Link Value</th>
<th>Max. Link Value In &amp; Out</th>
<th>Max Link Value In Only</th>
<th>Max Link Value Out Only</th>
<th>Remainder</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>3</td>
<td>42.9</td>
<td>128.7</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-3</td>
<td>1</td>
<td>14.3</td>
<td>14.3</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-4</td>
<td>2</td>
<td>28.6</td>
<td>57.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-10</td>
<td>1</td>
<td>14.3</td>
<td>14.3</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-1</td>
<td>5</td>
<td>100</td>
<td>300</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-4</td>
<td>1</td>
<td>100</td>
<td>100</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-5</td>
<td>1</td>
<td>33.3</td>
<td>33.3</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-6</td>
<td>1</td>
<td>33.3</td>
<td>33.3</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-9</td>
<td>1</td>
<td>33.3</td>
<td>33.3</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-6</td>
<td>1</td>
<td>100</td>
<td>100</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-7</td>
<td>1</td>
<td>20</td>
<td>20</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-8</td>
<td>1</td>
<td>20</td>
<td>20</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-1</td>
<td>5</td>
<td>60</td>
<td>180</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-6</td>
<td>1</td>
<td>100</td>
<td>100</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-1</td>
<td>1</td>
<td>100</td>
<td>100</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-6</td>
<td>1</td>
<td>100</td>
<td>100</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Total:** 1571.4
recorded in the second and third columns. In columns 5, 6, 7, and 8 of Table 5 check marks are entered to indicate whether each link value is: 1) the maximum value leaving a control and entering a display (Table 5, link 1-3); 2) the maximum value entering (Table 5, link 1-5); 3) the maximum value leaving (Table 5, link 5-6); or 4) none of the cases above (Table 5, link 1-4).

The information in Table 5 is used to generate a panel layout diagram in which controls and displays are oriented according to a sequencing principle/technique. Based upon this principle, displays and controls are arranged from left to right or top to bottom according to a series of rules described by Fowler, et al. (1958). A panel layout diagram for the 14E10 localization sub-task is shown in Figure 5. Solid lines indicate links which move from left to right in accordance with the sequencing principle. Broken lines indicate links which move left, directly up or down, or which move right but bypass one or more controls or displays. These latter links are in opposition to the sequencing principle and represent breaks in the operation sequence.

Three indices were derived from the data contained in Table 5. The first of these was the total number (N) of displays and controls used in performing a sub-task. For the 14E10 localization sub-task (Table 5), N equalled 10. The second index was the total number of response actions (R) comprising the sub-task. In the 14E10 localization sub-task R equalled 21. Finally, a total link value (LV) was obtained for each sub-task. In the example being used (Table 5), LV equalled 1534.4.

Based upon data contained in Figures 1 and 5 and in Table 5, five of the major indices described by Fowler, et al. (1968) were derived. The first of these (SE) expressed the degree to which the sequencing principle was applied to the console under consideration. It was calculated from the following formula:

\[
SE = \frac{\sum \text{link values}}{\text{total number of links}}
\]
\[ S^* = 2[100 \cdot \frac{OR_{\text{max}}}{OR_{\text{min}}} + \frac{100}{L_{\text{max}}} - 1] \text{, where} \]

where

- \( OR \): \( L \) runs total link values which are breaks in the operation sequence on the actual panel
- \( OR_{\text{max}} \): Same as above from the panel layout diagram (Figure 3)

1. Total number of sequencing lines on the actual panel.
2. \( L_{\text{max}} \): Total number of solid lines representing left-to-right links in the panel layout diagram (Figure 3).

For the 1P10 localization sub-task

\[ S^* = 2[100 \cdot \frac{086.6}{0[5, 5]} + 1190 \cdot \frac{0}{5}] = 50\% \text{.} \]

A second major index \( \Delta^2 \) reflected the extent to which the sub-task involved few or many alternative action choices out of each hardware item. The percentage of alternative actions present in a panel operation was given by the formula:

\[ \Delta^2 = 100\left(\frac{\Delta \cdot AA_{\text{min}}}{AA_{\text{max}} - AA_{\text{min}}} \right) \text{, where} \]

- \( \Delta \): \( AA \) is the total link value \( (\Delta) \) for an operation
- \( AA_{\text{max}} \): \( 100\%\) for \( \Delta \), where \( \Delta \) = total number of response actions
- \( AA_{\text{min}} \): \( 100\% \), where \( n \) = number of controls and displays.

In the 1P10 localization sub-task, \( \Delta = 21 \), \( n = 10 \), and

\[ \Delta^2 = \frac{1351}{1400} \cdot \frac{100}{10} = 50\% \text{.} \]

A third major index, breaks in operation sequence \( (BOS\%) \), expressed the degree to which it was necessary to employ controls and displays already used within the sequence of operations, or to use controls and displays employed predominantly during later stages of the operation.
sequence. These "reversals" or "jumps" represented breaks in the operation sequence and were reflected in the following formula:

\[ \text{BOS} = 100 \left( \frac{\text{BOS}_{\text{max}}}{\text{BOS}_{\text{max}}} \right) \]

where

\[ \text{BOS} = \text{the frequency of links which are "reversals" or "jumps"} \]

\[ \text{BOS}_{\text{max}} = TA - N + 1 \]

In the 14E10 localization sub-task

\[ \text{BOS} = 100 \left( \frac{\text{BOS}_{\text{max}}}{15} \right) = \% \]

A fourth index (\( F^5 \)) related to the relative frequency of use of display and control components. In essence it represented the degree to which the use of various components coincided with their arrangement in optimum reach envelopes. The index was given by:

\[ F = 100 \left( \frac{100F}{F_{\text{max}}} \right) \]

where

\[ F = \left| \frac{TA_1 - 1}{N/3} \right| + \left| \frac{TA_2 - 1}{N/3} \right| + \left| \frac{TA_3 - (\frac{2TA}{N} - 2)}{N/3} \right| \]

\[ F_{\text{max}} = \left| \frac{4TA}{N} - 4 \right| \]

\( TA_1, TA_2, \) and \( TA_3 \) = the total actions for each of three groups of controls and displays with each group containing one third of the total number (N). The procedure used to establish the groups is given by Fowler, et al. (1968, p. 22).

In the 14E10 localization sub-task

\[ F^5 = 100 \left( \frac{100 \cdot 10}{5,400} \right) = \% \]

55
A final index \( I_1 \) was derived from information shown in Table 2. This index indicated the degree to which important hardware components were located within different zones of the panel. The index was given by the formula:

\[
I_1 = 100\left(1 - \frac{100}{\max I^i}\right),
\]

where

\[
I^i = \left| \frac{1}{S} \sum I_i \right| = \left| \frac{X^i - X^{\max}}{X^{\max} - X^i} \right| + \left| \frac{Y^i - Y^{\max}}{Y^{\max} - Y^i} \right|
\]

\( I_i \), \( I^i \), \( I^\max \) = the number of controls and displays receiving importance ratings of three, two, or one (Table 2).

In the IO10 localization subtask,

\[
I_{10} = 100\left(1 - \frac{1}{15,5}\right) = 0^\circ.
\]

2.1.1.3 Additional Indices - Derivation of values for the remaining generic and specific indices was straightforward. Values for these indices were obtained primarily either from inspection of the types of data shown in Figures 1 and 3 and Tables 1 and 2, or during de-briefings with instructor personnel. As will be discussed later, however, not all of the required information (i.e., displayed/control ratios, signal/noise ratios, phosphor decay rates, etc.) was readily available.

2.2 Predictive Methodology

Generation of a set of quantitative indices for the description of trainee (or instructor) tasks is only the first step in resolving the training prescription-prediction issue. Given a set of indices, attention
must be given to their utility as predictors of learning rates or proficiency levels. The ability to make such predictions is, after all, the ultimate purpose in attempting to develop a quantitative task analysis methodology.

Use of the indices described above to predict learning rates or proficiency levels on actual Navy training devices was beyond the scope of the present effort. Nevertheless, it was felt that a demonstration of the predictive methodology would be valuable. Accordingly, a postdictive situation was arranged. A number of learning studies having a common performance measure were abstracted from the literature and described in terms of a selected set of quantitative indices. The purpose of this effort was to determine the feasibility of using quantitative task characteristic indices to predict learning rates or proficiency levels on different tasks.

Five major steps were involved in the postdictive exercise. These included: 1) development of a regression model; 2) selection of tasks; 3) selection of criterion measures; 4) selection and application of indices; and 5) data analysis.

2.2.1 Regression Model

A multiple-regression model was developed in which task characteristic indices were treated as predictor variables. The model was based upon the premise that the indices could be used to predict average learning rates or proficiency levels on different tasks. There were two restrictions on the model. First, tasks had to be described in terms of the same set of indices. Second, tasks had to share a common response measure (e.g., time on target, probability of detection, etc.). The rationale involved in generation of the model is presented below.

Postdictive simply refers to the fact that existing criterion data was used, whereas in prediction, arrangements are made to collect data in accordance with some specific experimental design. Postdictive entails precise control over many variables in order to enable complete replication set of data for analysis.
In the conventional prediction problem, subject characteristics (e.g., aptitude or ability test scores) are treated as predictor variables and are used to predict how individuals in a particular group will perform on a specific task. This leads to a multiple-regression equation which can be represented as:

$$Y_{i1} = \beta_{10} + \beta_{11}X_{i1} + \beta_{12}X_{i2} + \cdots + \beta_{1k}X_{ik}$$  \hspace{1cm} (Eq. 1)

where

- $Y_{i1}$ = predicted performance for individual "i" on task 1
- $\beta_{1k}$ = regression coefficient for the kth predictor variable
- $X_{ik}$ = score of individual "i" on predictor variable "k".

If the subject predictors in Eq. 1 were all standardized variables while the measures $Y_{i1}$ were not, then $\beta_{10}$ would equal $Y_{i1}$, the mean of the observed criterion measures. In Z-score form the regression equation could be:

$$Y_{i1} = \bar{Y}_{1} + \beta_{11}Z_{i1} + \beta_{12}Z_{i2} + \cdots + \beta_{1k}Z_{ik}$$  \hspace{1cm} (Eq. 2)

Equation 1 or 2 would provide information on the relevance of various subject characteristics for predicting the performance of an individual on task "1". But task "1" has its own characteristics. These characteristics are fixed. That is, they are constants which are not represented in Eq. 1 or Eq. 2, and as such they cannot influence the predictor coefficients ($\beta_{1k}$).

Suppose, however, that the original group of individuals performs on a second task (2). In addition to Eq. 2 there will now be another multiple-regression equation:

$$Y_{i2} = \bar{Y}_{2} + \beta_{21}Z_{i1} + \beta_{22}Z_{i2} + \cdots + \beta_{2k}Z_{ik}$$  \hspace{1cm} (Eq. 3)
If the two equations (Eq. 2 and Eq. 3) differ only with respect to the first coefficient ($\hat{\gamma}_1$ and $\hat{\gamma}_2$), and the tasks have been measured in terms of some common performance metric, then this is equivalent to finding a significant difference between means. The difference between these can only be "explained" in terms of differences between the tasks themselves. (In the present study the attempt has been made to represent these differences quantitatively, in terms of the task characteristic indices.)

If the concept of differences between tasks and consequent differences between means is extended to a set of tasks, performed by the same group of individuals, a variable ($\hat{\gamma}_m$) is created. It can be hypothesized that the specific values of this variable should be predictable in terms of task characteristic indices. The multiple-regression equation would have the following form:

$$\hat{\gamma}_m = b_{0m} + b_{1m} l_1 + b_{2m} l_2 + \ldots + b_{nm} l_n$$

where

- $\hat{\gamma}_m$ = predicted mean performance score on task "m" (e.g., $\hat{\gamma}_1$ in Eq. 2, or $\hat{\gamma}_2$ in Eq. 3, etc.)
- $b_{nm}$ = regression coefficient for the nth task index
- $l_{mn}$ = index value for task "m" on task index "n".

Equation 4, therefore, was explored during the post-dictive study.

2.2.2 Selection of Tasks

Four criteria were established to aid in the selection of studies which could be used in the post-dictive exercise. The first criterion arose from the need to express the performance measures in terms of
common metric. Consequently, only those studies were considered which reported a "percent time on target" performance measure. These were studies in which various tracking tasks were employed. The second criterion required that a learning curve be reported in which data were available for at least ten different points in time. A third criterion was that each data point in the learning curve be based on a minimum of 10 observations. The fourth and final criterion was that there could be no change in experimental conditions over the course of the learning curve (e.g., administration of a drug after the fifth session, etc.).

Approximately 950 studies were examined in terms of the four criteria. From this sample a set of only 22 studies met all four criteria! These studies represented the following kinds and numbers of simple laboratory tracking tasks:

- a) rotary pursuit (6);
- b) two-hand coordination (4);
- c) pedestal sight manipulation (5);
- d) specialized tracking tasks (3);
- e) rudder control (2);
- f) turret pursuit (1);
- g) pilot simulator (1);
- h) Iowa pursuit apparatus (1);
- i) wheel turning (1).

This sample was used as a basis for generating the desired regression equation. The specific studies which were employed are referenced in Appendix D.
2.2.3 Criterion Measures

The basic learning data for each of the 22 studies consisted of the percent time on target (TOT) attained in relation to the number of minutes of practice on given tasks. Three types of criterion measures were derived from these data including:

1. the percent time-on-target attained after 1.25, 5, 10, 15 and 20 minutes of practice;
2. the amount of practice (expressed in minutes) required to reach TOT proficiency levels of 41", 51", 61", and 71%; and
3. the increments in TOT after 10, 15, and 20 minutes of practice against a TOT base-line at five minutes (e.g., TOT at 10 minutes minus TOT at five minutes, etc.)

The first two sets of criterion measures addressed themselves to the effects of practice on proficiency levels. The third set was employed to obtain measures reflecting rates of learning. These measures were adjusted in terms of an early level of proficiency in an attempt to equate the different subject samples on early skill levels.

Unfortunately, not all twelve criterion measures were not available for each of the 22 tasks sampled. For example, subjects in some studies failed to reach 71% TOT; in some studies 20 minutes of practice was not given. The net result of this attrition was that different numbers of studies were available for different criterion measures. Because the sample was small to begin with, criterion measures were employed for which a minimum of 18 studies was available. The three criterion measures finally employed were: 1) TOT after five minutes of practice; 2) the time (minutes) required to reach a TOT level of 41%; and 3) the increment in TOT between five and 15 minutes of practice.
2.2.1 Application of indices

The twenty-two tracking studies were obtained in their original forms from the literature; when references were given to more complete descriptions of apparatus these were also acquired. Each tracking task was described in terms of its task characteristic rating scales. The scales were applied by two trained raters who worked independently.

Upon completion of the ratings, the results were compared and an average set of ratings was derived for each tracking task under the following rules. The two ratings on each scale were averaged if they were no more than two points apart. If they differed by more than two points, they were discussed in the presence of a third rater who served as a final arbiter. Approximately 25 percent of the 586 ratings (i.e., 18 indices applied to 22 tasks) required arbitration. It should be noted, parenthetically, that five more scales were applied to the 22 tracking studies than were used to describe the trainee sub-tasks in the sonar training devices. This increase resulted from the use of certain indices (e.g., degree of muscular effort involved) which seemed germane to the variety of tasks sampled from the literature but which did not seem relevant to surveillance system sub-tasks.

The DII and major panel lay-out indices could not be applied to the types of simple, laboratory tracking tasks taken from the literature. Nevertheless, an additional set of 10 indices was also applied to the tracking tasks including such descriptors as: 1) number of tracking dimensions; 2) number of control elements; and 3) number of display bits. Consequently, each of the 22 tasks was described in terms of 28 indices.

In regression analysis, however, the number of predictors should never approach, let alone exceed, the number of cases sampled. As the number of predictors approaches the number of cases sampled, the multiple-regression coefficient becomes spuriously large and uninterpretable. In
the present study, therefore, five rating-scale indices were selected as the basis for one analysis, and five "human-engineering" indices were used in another analysis. Indices were selected from the total set of 28 on three bases: 1) hypothesized relevance to the criterion measures; 2) relatively low intercorrelations with other indices; and 3) normality of distribution.

The five rating-scale indices included:

a) degree of muscular effort involved (MUSC);
b) simultaneity of responses (SIMI);
c) number of output units (UNIT);
d) number of responses (NO, R); and
e) variability in stimulus location (VARS).

The five additional indices selected for separate analysis included:

a) number of tracking dimensions (TRAC);
b) number of control elements (CONU);
c) number of display digits (DISI);
d) length of duty cycle (DUTY); and
e) ratio of practice time to duty cycle (DUTR).

2.2.5 Data Analysis

Six multiple-regression analyses were conducted by evaluating the two independent sets of predictors in terms of the three criteria specified above. Three additional analyses were also conducted in which predictors from the two different sets were combined. The multiple-regression coefficients were generated by performing linear step-wise regression analyses. All analyses were conducted on an IBM 1130 computer using standard statistical programs.
2.5 RESULTS AND DISCUSSION

Two different sets of results are described and discussed in this section. The first set of results concerns the attempt to apply quantitatively indices to information obtained from an analysis of trainee data in some training devices. The second set relates to the post-training studies that were undertaken.

5.1.1.1 Task Information

Task-specific information was derived in order to apply two different sets of quantitative indices. Generic indices were used which cut across trainee sub-tasks. They were comprised of DFI, panel layout and task type, rating-scale, and miscellaneous indices. The second set of indices were more specific. They were relevant to some but not to all of the training sub-tasks.

5.1.1.1.1 Generic Indices

End-analytic data obtained from some training devices were used to derive the "human-engineering" and miscellaneous index values summarized in Table 5 and the rating-scale index values shown in Table 6. In both tables, index values are presented for each trainee sub-task and device. Mean values for each sub-task are also shown to facilitate examination of the data.

5.1.1.1.1.1 Display by Learning Index - As shown in Table 5 the DFI appears to discriminate well both between and within trainee sub-tasks. Lowest information transfer (smallest DFI values) is shown for the set-up and classification sub-tasks. The relatively low values for
Table 6
Generic "Human Engineering" Indices Derived From Task Analysis Data

<table>
<thead>
<tr>
<th>Sub-Task</th>
<th>Device</th>
<th>DEI</th>
<th>10^-3</th>
<th>TA</th>
<th>L</th>
<th>LV</th>
<th>DC</th>
<th>AC</th>
<th>HI</th>
<th>MI</th>
<th>PM</th>
<th>ACER</th>
<th>PERC</th>
<th>WARS</th>
<th>GRRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coils A</td>
<td>1G3 (Orig.)</td>
<td>33.3</td>
<td>30</td>
<td>22</td>
<td>2199.9</td>
<td>43</td>
<td>97</td>
<td>17</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coils A</td>
<td>1G3 (Revised)</td>
<td>35.7</td>
<td>24</td>
<td>28</td>
<td>2200.2</td>
<td>44</td>
<td>97</td>
<td>18</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14120</td>
<td>15.9</td>
<td>27</td>
<td>25</td>
<td>2899.9</td>
<td>56</td>
<td>47</td>
<td>27</td>
<td>10</td>
<td>14</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antenna Sym</td>
<td>14120</td>
<td>7.2</td>
<td>37</td>
<td>22</td>
<td>2299.9</td>
<td>32</td>
<td>67</td>
<td>27</td>
<td>23</td>
<td>15</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>24</td>
<td>32</td>
<td>24</td>
<td>2100.0</td>
<td>33</td>
<td>25</td>
<td>75</td>
<td>19</td>
<td>10</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Patent 5000 | 1G3    | 17.7   | 9    | 4 | 500   | 47 | 40 | 37.6 | 3  | 3  | 2    | 0    |      |      |
| Patent 5000 | 1G30   | 31.2   | 14   | 2 | 699.7 | 44 | 33 | 47.4 | 8  | 12 | 3    | 3    |      |      |
| Patent 5000 | 1G30-2  | 28.8   | 14   | 5 | 32.6  | 13 | 32.5 | 43 | 4  | 3  | 3    |      |      |      |
| Mean       | 33     | 12    | 5  | 573   | 14 | 57 | 32.6 | 4  | 2  | 3    |      |      |      |

| High Power  | 14120  | 65.6   | 7    | 5  | 500.0  | 37 | 43 | 37.6 | 3  | 2  | 2    | 0    |      |      |
| High Power  | 14120-2 | 34.5   | 23   | 11 | 225.6 | 50 | 7  | 47.4 | 7  | 7  | 2    |      |      |      |
| Mean       | 47     | 10    | 3  | 129.3 | 16 | 34 | 37.6 | 3  | 2  | 3    |      |      |      |

| Low Power  | 14120  | 3.5    | 4    | 7 | 22.5  | 51 | 77 | 55.4 | 9  | 9  | 7    |      |      |      |
| Low Power  | 14120-2 | 5.8    | 7    | 6 | 25.4  | 50 | 77 | 55.4 | 9  | 9  | 7    |      |      |      |
| Mean       | 12     | 15    | 6  | 25.3  | 15 | 77 | 55.4 | 9  | 9  | 7    |      |      |      |
classification reflect the poor classification performance which has been reported both for passive and active sonar surveillance (Levy and Minchella, 1963). This result lends credence to the diagnostic value of the DEI. Further credence is obtained from a comparison of the values derived for an original and a revised set-up procedure on the ITI. The original sequence given to the investigators by the ITI instructor included steps which appeared to be inappropriate (i.e., using controls instead of indicators for feedback) and unduly repetitive. Therefore, the procedure was revised slightly, and the DEI recalculated. The outcome was an improvement in information transmission.

The DEI requires detailed task-analytic information for its extraction. Given that information, however, extraction of the DEI is for the most part routine. The major difficulty encountered was determination of the number of states which controls, and in particular, displays could assume. An example of this problem was determination of the number of display states for various uses of the ITI. To deal with this problem a number of conventions were adopted which were consistent from analysis to analysis. Other analysts such as Siegel, however, possibly might have applied other interpretations. For example, when the ITI was used for tracking, a nine-state display was assumed for both bearing and range controls. Whether the operator was manipulating his bearing or range hand-wheel he had to "read" cursor position vis-a-vis the target in one of nine basic locations (e.g., on bearing, leading, lagging; on range, short, long; and two-way combinations of these).

In general, however, application of the DEI was straight-forward. Values could be obtained fairly quickly, reliability did not appear to be a problem, and the index differentiated sub-tasks and devices. The
DEI possessed diagnostic value and was intuitively satisfying, varying in accordance with subjective impressions of sub-task difficulty.

3.1.1.2 Panel Lay-Out and Task-Type Indices - The seven panel lay-out indices shown in Table 6 also differentiated between and within sub-tasks. The largest number of responses (TA) and control and display components (X) were found in the set-up sub-tasks, while these same sub-tasks also possessed the largest link values (LI). Of interest was the fact that fairly large link values were also found for localization and classification, sub-tasks involving few displays and controls. In these cases the large link values reflected the cyclic or repetitive nature of performance. Fewer controls and displays were utilized than during set-up, but they were employed more frequently.

Values obtained for the sequencing technique index ($S^*$) can vary from zero to 100. The higher values on this index reflect relatively better panel lay-outs in terms of certain sequencing principles which are intended to enhance operator performance. With this interpretation in mind, two features of the $S^*$ data shown in Table 6 are of interest. First, within each of the three devices, as one moves from set-up through detection, localization, and classification, the $S^*$ values increase successively. These data suggest that the sonar stacks have been designed to facilitate classification performance, at least in terms of panel lay-out, with compromise designs being employed in the other sub-tasks. Second, the $S^*$ values associated with the original and revised IHEIS set-up sub-tasks appear to be inverted (i.e., the higher value might have been expected for the revised procedure). There is no inconsistency however. The data simply suggest that if one is to revise an inefficient procedure, one must also revise the panel lay-out upon which that procedure is based.
Interpretation of the "time in operation sequence" (TOS), "alternative action" (AA), and "vulnerability of use" (VU) indices is difficult. The indices are designed to measure and describe the importance of use at different levels of the task. Differentiation and increased data handling were used to increase the analysis. These particular indices, because of analysis to some extent more labile than others, valuing score variability can lead to significant changes. These indices can become sensitive to changes in data collection procedures involving feedback. The analysis required, however, before this point can be resolved.

The panel layout indices require the same type of detailed test-analytic index as used in the TOS. The panel layout indices which are based on percentages, however, are much more difficult to generate than the TOS, and are conceivably less reliable. The basic problem lies in translating the test-analytic data into the panel layout data in the way of several of the indices. In this study "ties" between the TOS indices associated with different links were found repeatedly (Chidley & Dowler, et al., 1963). It has not been determined the number of conventions were adopted in the present study to resolve this problem, but the conventions involved a greater degree of arbitrariness than seemed desirable. More rigorous rules need to be developed in this area before the indices can be used with complete confidence.
3.1.1.3 Miscellaneous Generic Indices - The third group of indices shown in Table 6 represented an attempt to analyze the total number of trainee actions (TA) into more specific types. These included:

a) control responses (CRs) which were direct manipulations of controls in response to a display reading, fixed instructions, or to the instructor’s directions;

b) feedback responses (FBs) which represented display “readings” to corroborate the effects of control actions;

c) information acquisition responses (IARs) which represented readings of displays to acquire information; and

d) instructor initiated responses (IRs) which represented cases when an instructor entered or told trainees to enter a specific value into a control.

In addition to these response-related indices, the number of non-normal repertoire responses (NNRAs), the number of redundant information sources (RIs) and the time for sub-task completion (Time) were also ascertained for sub-tasks in each device. These three indices are not shown in Table 6 because they failed to differentiate at all. The four indices (i.e., a) to d) above) which are included in Table 6 did not differentiate as clearly among the sub-tasks as did others previously described. Nevertheless, they were easy to apply and generated data of some interest. For example, instructor initiated responses were most evident among the set-up sub-tasks. The value fell to zero for localization and classification. Similarly, the information acquisition responses varied noticeably both within and between sub-tasks.
3.1.1.1 Task Characteristic Rating Scales - Ten of the thirteen scales which were applied to the task analysis data are shown in Table 7. The three scales which are not shown were modified to provide enumerative rather than rating data. These scales correspond to the V, TV, and TM indices previously discussed in Table 6.

The most striking feature of the data presented in Table 7 is the similarity of index values across sub-tasks. Although some differentiation both within and between sub-tasks is obtained, it is not as pronounced as that shown in Table 6. This, of course, is due to the fact that most of the ratings were based upon seven-point scales. Consequently, the range of possible index variation was very restricted relative to the ranges possible for other indices. The restricted range of values which the scales can assure is not necessarily a liability. Even within this small range different scale values may actually reflect differences in performance.

Insofar as possible, however, the rating scales employed in the present study should be revised. Many of the constructs upon which the scales are based seem valuable. However, other means for their quantification should be considered. More direct measurement of these task characteristics would not only enhance index reliability, but would also permit a greater range of variation. These modifications would result in indices which are sensitive to differences among fairly complex tasks, differences which at present are possibly being minimized.

3.1.2 Specific Indices

In addition to the various generic indices described above, an attempt was made to apply a set of 15 specific and 10 training technique indices. The results were generally inconclusive (i.e., many specific indices could not be applied; when they could be, they did not clearly discriminate among tasks; and training indices were simply binary statements about the presence or absence of a "fuzzy" capability, for instance). Indices which could not be readily applied included range in signal-to-noise ratio (SN), hearing and range control-display ratios, and signal persistency. SN was directly manipulable only on the TV. The
Table 7
Task Characteristic Ratings Derived
From Task Analysis Data

<table>
<thead>
<tr>
<th>Sub-Task</th>
<th>Device</th>
<th>Task Characteristic Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>UNPL</td>
</tr>
<tr>
<td>Setup</td>
<td>1983 (orig.)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1983 (rev.)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1986</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1986-266X</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>1</td>
</tr>
<tr>
<td>Detection</td>
<td>1983</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1986-266X</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>1</td>
</tr>
<tr>
<td>Localization</td>
<td>1983</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1986-266X</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>1</td>
</tr>
<tr>
<td>Classification</td>
<td>1983</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1986-266X</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>1</td>
</tr>
</tbody>
</table>
other devices, using tape inputs, made use of unspecified ratios. Control-display ratios presented a problem since they varied with range scale and therefore no single value could be obtained. The persistency index required information about phasor decay rates which was not readily available. However, all these indices are potentially available to a system designer and they are significant for the training process (Corcoran, Carpenter, Webster & Woodhead, 1968; "Indie & Pumphian, 1964; "Irabella, 1969; "Nichols & Cotterman, 1958; "Short & Hargrey, 1967). Therefore future attempts should be made to include them in a predictive scheme.

The remaining 11 specific indices (see pp. 17-18) could be applied but appeared to be of limited value. Some such as variation in target range and speed and sequencing of problems were applicable across detection, classification and localization sub-tasks. These indices provided for little discrimination among devices and for no discrimination within a device (i.e., problem characteristics were not varied across sub-tasks within devices). Other specific indices were applicable to only one of the four sub-tasks. Some, such as bearing and range error tolerance and especially number of tracking dimensions did not discriminate well. Others did vary among devices. These included: number of cues available for classification, number of cues used simultaneously, number of false targets used, target to non-target ratio, and number of contacts per minute. Values for these indices were fairly unstable, however, since they were merely rough estimates, given by instructors, of the materials appearing on various tapes. Finally, information about different training techniques was of little value for the purpose of the present study, because they provided no differentiation among sub-tasks within devices, and little if any discrimination among the devices sampled.

With few exceptions, the specific task indices appeared to be arbitrarily chosen or fixed by equipment design. The lack of flexibility, in the taped devices particularly, was emphasized to the investigators.

3.2 Prediction Studies

3.2.1 Rating Scales

Five task characteristic rating scales were used to predict the criterion of percent-time-on-target (PTO) after five minutes of...
<table>
<thead>
<tr>
<th>Analysis</th>
<th>Predictors</th>
<th>Criterion</th>
<th>No.</th>
<th>No. Pred.</th>
<th>R</th>
<th>R²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>UNIT SIGH NO.R DISC VARS</td>
<td>% TOT at 5-min.</td>
<td>20</td>
<td>5</td>
<td>.78</td>
<td>.60</td>
<td>.00</td>
</tr>
<tr>
<td>2</td>
<td>Same as #1</td>
<td>41% TOT 18</td>
<td>4</td>
<td>.66</td>
<td>.44</td>
<td>.10</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Same as #1</td>
<td>Adj.15-min. 18</td>
<td>5</td>
<td>.55</td>
<td>.31</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>TRAC CORE DISO DUTY WORE</td>
<td>% TOT at 5-min.</td>
<td>20</td>
<td>4</td>
<td>.61</td>
<td>.40</td>
<td>.10</td>
</tr>
<tr>
<td>5</td>
<td>Same as #4</td>
<td>41% TOT 18</td>
<td>4</td>
<td>.68</td>
<td>.46</td>
<td>.10</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Same as #4</td>
<td>Adj.15-min. 18</td>
<td>5</td>
<td>.48</td>
<td>.22</td>
<td>.10</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>UNIT NO.R VARS TRAC CORE</td>
<td>% TOT at 5-min.</td>
<td>20</td>
<td>5</td>
<td>.62</td>
<td>.63</td>
<td>.05</td>
</tr>
<tr>
<td>8</td>
<td>NO.R VARS DISO DUTY WORE</td>
<td>41% TOT 18</td>
<td>5</td>
<td>.78</td>
<td>.61</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>SIGH NO.R VARS CORE DUTY</td>
<td>Adj.15-min.</td>
<td>18</td>
<td>5</td>
<td>.57</td>
<td>.33</td>
<td>.05</td>
</tr>
<tr>
<td>10</td>
<td>Same as #7</td>
<td>% TOT 10-min.</td>
<td>10</td>
<td>5</td>
<td>.74</td>
<td>.55</td>
<td>.05</td>
</tr>
</tbody>
</table>
Table 8 (Continued)

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Predictors</th>
<th>Criterion</th>
<th>No. Cases</th>
<th>No. Pred. Used</th>
<th>R</th>
<th>R²</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Same as TOT</td>
<td>15-min.</td>
<td>20</td>
<td>5</td>
<td>.64</td>
<td>.41</td>
<td>n.s.</td>
</tr>
<tr>
<td>12</td>
<td>NO, P</td>
<td></td>
<td></td>
<td></td>
<td>.63</td>
<td>.40</td>
<td>&lt;.10</td>
</tr>
<tr>
<td></td>
<td>VARS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DISO</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DUTY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Same as TOT</td>
<td>10-min.</td>
<td>19</td>
<td>5</td>
<td>.71</td>
<td>.51</td>
<td>&lt;.10</td>
</tr>
<tr>
<td>14</td>
<td>Same as TOT</td>
<td>15-min.</td>
<td>20</td>
<td>4</td>
<td>.69</td>
<td>.48</td>
<td>&lt;.05</td>
</tr>
</tbody>
</table>

ONUT: number of output units
SIMH: simultaneity of responses
NO, P: number of responses
MPSC: degree of muscular effort involved
VARS: variability in stimulus location
TRAC: number of tracking dimensions
CONE: number of control elements
DISO: number of display digits
DUTY: length of duty cycle
WORK: ratio of practice time to duty cycle

54
practice. The results, shown in Table 8 as Analysis 1, indicate a multiple $R$ of 0.78 which accounts for 60% of the variance in the criterion measure; the correlation was significant beyond the .05 level of confidence. This criterion was viewed as the initial level of proficiency (an earlier proficiency level obtained at 1.25 minutes was not used due to the smaller number of studies yielding this measure).

These same five rating scales were then entered into regression analyses to predict two other criterion measures: 1) the time required to reach 41% TOT, and 2) an adjusted % TOT achieved after 15 minutes of practice. The latter measure was derived by subtracting the % TOT at five minutes from that reached at 15 minutes, i.e., it reflected the increment in % TOT over the initial proficiency level. As indicated in Table 8, neither of these analyses ($R^2$, 0.5) yielded a significant multiple $R$ ($p > .10$). A tentative hypothesis was formed which suggested that the task characteristic scales had maximum predictive efficiency for initial levels of performance rather than performance levels achieved later in training.

### 3.2.2 "Human Engineering" Indices

Analyses aimed at predicting these same three criterion measures were then conducted using five "human-engineering" indices. These indices, shown in Table 8 as analyses 4, 5, and 6, yielded two multiple $R$'s which were significant at a less stringent level of confidence ($p < .10$); the 0.10 confidence level was considered appropriate for this exploratory work. The two criterion measures successively predicted were the initial level of performance (% TOT at five minutes) and the time required to reach the 41% TOT level. Contained within the set of "human engineering" indices were two predictors, duty-cycle length and work ratio, which reflected the distribution of practice dimension inherent in all of the studies. An examination of their contributions
to the predicted variance indicated that they were minimally involved in predicting initial performances but made increasing contributions to prediction of the later performance level of H, DT. These findings are typical in that we could not expect predictors representing a distribution of practice variable to be as effective in predicting earlier performance as the wider be later in the course of training. The remaining three indices had their highest predictive efficiency in regard to the initial performance criterion.

In general then, the five task-characteristic scales and three of the "human engineering" indices performed best in predicting starting levels of performance; increases in predictive efficiency for later performance were noted for the two predictors representing distribution of practice.

The results from the analyses up to this point are understandable on a post-hoc basis when it is recognized that the rating scales and "human engineering" indices are mainly descriptive of the task per se and not of the training regimen or conditions of learning under which the task was performed. The training variables were minimally reflected in the two predictors representing the distribution of practice dimension; these nevertheless did show increasing contributions to prediction when a criterion involving later performance was used.

A.2.4 Combined Indices

The next step taken in the analysis was to create a new set of predictors based on the "best" predictors from the task-characteristic scales and the "human engineering" indices. The composition of this combined set varied in accordance with the criterion being predicted. Table 8 shows which criterion was used for each of the three criterion measures.
The net effect of combining the scales and indices was to increase the multiple R's obtained for the initial performance level and the 41% TOT criteria. Multiple correlations of 0.82 (p < .01) and 0.78 (p < .05) were obtained, respectively. The combined set of predictors yielded no increase in prediction of the criterion measure of the adjusted 15-minute % TOT nor was the multiple-correlation significant (p > .10). These analyses are represented in Table 8 as 7, 8, and 9.

At this point in the analysis it was decided to investigate further the hypothesis that the rating scales and indices were primarily predictive of initial rather than interim performance. To test this point two additional criterion measures, % TOT at ten and 15 minutes, were used. The predictors employed were the combined set which had been maximally effective for the five-minute initial level of performance (see analysis 7 in Table 8). The result of these two regression analyses (10 and 11 in Table 8), lent support to the hypothesis in that the multiple correlations decreased as the length of practice represented by the criterion increased. These relationships are shown below.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Multiple R</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>% TOT 5-min.</td>
<td>.82</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>% TOT 10-min.</td>
<td>.74</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>% TOT 15-min.</td>
<td>.64</td>
<td>ns</td>
</tr>
</tbody>
</table>

Having demonstrated, within the limits of the study, that the majority of the scales and indices had their maximal predictive efficiency for the starting level of performance, one further set of analyses was conducted to test the prediction limits for interim performance. In these analyses (12, 13, and 14 in Table 8), the set of predictors used were those previously employed to predict 41% TOT (analysis 8). Contained in this set were the distribution of practice.
predictors. Using the "best" scales and indices plus the duty-cycle
and work-ratio predictors, regression analyses were run with the five,
ten and 15-minute TOF criteria. The results, shown below, indicate
that this combination of predictors effectively extended the pre-
dictive efficiency to interim levels of performance:

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Multiple R</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. TOF 5-min.</td>
<td>.65</td>
<td>.10</td>
</tr>
<tr>
<td>2. TOF 10 min.</td>
<td>.71</td>
<td>.10</td>
</tr>
<tr>
<td>3. TOF 15-min.</td>
<td>.69</td>
<td>.05</td>
</tr>
</tbody>
</table>

An examination of the individual predictors across practice time
(i.e., five, ten, and 15-minutes) again indicated that duty-cycle and
especially work-ratio made increasingly greater contributions to the
amount of predicted variance as practice time grew longer. A similar
increasing contribution was also found for the "number of responses"
predictor; a recheck of this predictor's contribution across practice
time was made in analyses 10 and 11 and a similar profile appeared.
Thus, it would seem that at least three of the indices are capable of
extending predictive efficiency to the interim levels of performance.

3.2.4 Discussion

An overall appraisal of the findings of the 11 regression analyses
indicates, first, that the criterion measures used in the post-iction
studies are of two distinct types. Initial level of performance, the
first type, appears to be predicted most efficiently by descriptors
which relate to features of the task per se. The majority of the scales
and indices are directed to just this point. At initial levels of per-
formance, the training variables used in the studies have had little if
any impact. Dominant factors at this stage are probably aspects of the
task itself and the abilities of the subjects.

This "lower" correlation has a higher n-value due to the slightly
larger number of cases contributing to it as compared to the R of .71
directly above it.

58
It is conceivable that the bulk of the residual variance in the initial performance predictions resides in the subject factor. Several studies (Fleishman 1957, 1960; Fleishman and Hempel, 1951, 1955; and Hinrichs, 1970) have shown that the abilities of subjects are related to or predictive of performance on a variety of tasks. These studies have also indicated, however, that the pattern of abilities contributing to proficiency on complex tasks may change as practice on such tasks continues and proficiency increases.

When attention is focused on predicting interim performance levels, the second type of criterion measure, predictive efficiency of the majority of the predictors declines. One potential explanation for this decline would be the increasing impact of whatever training variables are in effect plus the interaction of these variables with subject characteristics. The analyses indicate that changes in the predictive contributions of the various indices occur when the criterion measures reflect interim rather than initial performance. A smaller number of the predictors appear to come into their own when later performance is examined. Understandably, some of these predictors relate to the important training dimension of distribution of practice, i.e., massed versus spaced practice. In a less readily understood case, the predictor variable of "number of responses" also increased in predictive efficiency as practice time increased.

Had other types of task-characteristic indices been used in the post-diction study (e.g., DH1 and panel indices), their predictive efficiency might have increased as proficiency increased. If obtained, this finding would be of interest when compared with studies on the contribution of subject variables (abilities) to different stages of practice (e.g., Fleishman, 1960; Hinrichs, 1970). These and similar studies have generally shown decreasing predictions from ability variables with continued practice and higher proficiency levels on the criterion tasks. At the same time they have identified "task-specific" factors which increase.
in importance as proficiency increases. An interesting area of research would be to determine whether the "task-specific" factors found in such studies can be related to or explained in terms of task-characteristic indices. The simultaneous use of subject, task, and training indices might enhance predictive efficiency.

Granting all of the limitations inherent in the post-diction study, its results confirm the initial conceptualization of the training situation which viewed it in terms of the task per se, the subjects, the training variables, and interaction among these components. The evidence at hand indicates the need for indices specifically designed to measure aspects of each of these components.

The problems and limitations of the post-diction are many and should not be slighted. The attrition experienced as the search went on for suitable studies in the open literature placed a decided limitation on how far the results of the regression analyses may be generalized. The small number of studies acquired did not permit the important step of cross validation to be taken. Use of the literature itself removed any control over the subject factor and prevented application of the DEI and panel lay-out indices. Yet in spite of these factors, sizeable portions of the total variance in performance were predictable using the indices applied. Shrinkage of these figures would undoubtedly occur upon cross validation but their continued development still appears justified.

The development should proceed, in three directions. First, refinement of the rating scales must be undertaken. Several instances were encountered during both the field work and the post-dictive studies in which inter-rater agreements were lacking. Agreement can be improved by providing more concise and consistent definitions for the rating-scale indices. Agreement would also be improved to the extent that the concepts represented by the scales can be measured or enumerated more directly.
Second, attention must be given to development of training technique indices. Results of the post-diction exercise suggest that such indices may aid in the prediction of advanced levels of proficiency. Whether these indices would be as necessary in dealing with complex training devices as they appear to be when studying simple laboratory tasks remains to be seen. Third, and finally, the types of indices employed in the present study must be applied to actual training devices for which performance criteria are available. The demonstration of relationships between quantitative task indices and performance measures in highly complex man-machine tasks would be truly impressive.
1.0 CONCLUSIONS

This study has demonstrated the feasibility of using a variety of quantitative indices to describe salient characteristics of the trainee sub-tasks found in surveillance system training devices. Although applied only to trainee sub-tasks in selected sonar systems, many of the indices nevertheless appear applicable to a wide variety of both trainee and instructor tasks.

The importance of this demonstration is evident when one considers the nature of many of the quantitative indices which were employed. First, several of the measures, and in particular the BEI and panel lay-out indices, are directly related to features of a task familiar to design engineers. These are hardware and procedural features which might be reconfigured during the development of alternative designs. Modifications of these task characteristics would be reflected by changes in the values of many of the quantitative task indices employed in the present study. Second, and more importantly, these same task characteristics can be hypothesized to bear a relationship to measures of task performance including proficiency levels or rates of skill acquisition. Fowler, et al. (1968) have already demonstrated this type of relationship for some of their panel lay-out indices.

In theory, therefore, the possibility exists of developing quantitative profiles of tasks and of relating such profiles to measures of performance. Were information of this type available, it would then be possible to predict the behavioral consequence of restructuring a task's profile of quantitative indices. A basis would exist for predicting the effectiveness of alternative training device designs. All of this is contingent, of course, upon the demonstration of a relationship between the quantitative indices and measures of performance.
Results of the post-diction study conducted during this project confirmed the existence of significant relationships between "task characteristic indices" and measures of performance. The relationships were strongest (p < .05) during early stages of practice. However, nine of the 14 multiple-correlation coefficients which were computed suggested the presence of relationships between task indices and criterion measures of performance (p < .10). These results were particularly encouraging, being obtained in spite of the fact that the major indices of interest (DHII and panel lay-out indices) could not be employed, and that differences between groups of subjects (a violation of the predictive model) could not be avoided.

The major conclusion of this study is that further development of a quantitative task analytic methodology is warranted. If the types of indices employed in this study can be related to behavioral measures obtained in training devices, an invaluable tool can be developed for individuals responsible for sound training decisions. At the very least this approach would put the device design process on a more objective footing.
5.0 RECOMMENDATIONS

As has been discussed in earlier sections of this report, development of a set of quantitative indices for the description of tasks is only the first step in a larger program of research. To be of applied value it must be demonstrated that changes in the design of a training device (i.e., change in the trainee's or instructor's station and tasks) are reflected in corresponding quantitative task indices. Similarly, it must be shown that variations in the quantitative indices are related to different rates of learning or levels of proficiency.

Additional research will be required to demonstrate these relationships. For a number of reasons, however, the investigation will be difficult to execute satisfactorily. These problems stem from the requirement that the research focus on actual training devices currently in use in the field.

The major recommendation stemming from the present report is that a predictive study be undertaken, based on actual training devices. As the first step in this study, additional quantitative indices should be assembled. Emphasis should be placed upon use of indices previously developed and reported in the literature, rather than on the development of new indices. Most desirable would be indices possessing the following attributes: 1) construct validity; 2) ease of quantification and reliability; 3) generic applicability across devices, and 4) based upon task features of relevance to design engineers.

The assembled indices should be applied to a large sample of both trainee and instructor tasks for which criterion data are available.
In this future effort consideration should be given to use of a wider variety of criterion measures. In addition to rate of learning and level of proficiency, consideration should be given to transfer of training criteria. Similarly, measures of instructor proficiency during device operation must be entertained.

If the results of these efforts were promising, then a number of interactive studies should be undertaken. These studies would attempt to generate guidelines about the personnel who would benefit most from training, and about the training techniques which could be applied to particular tasks to increase training effectiveness. The key to this research, however, is to first understand the relationships between different tasks and the performance of those tasks.
REFERENCES


Smith, B. J. Task analysis methods compared for application to training equipment. NTDC 1218-5, (September 1965). H. S. Naval Training Device Center, Port Washington, N. Y.

Stolourow, I. A taxonomy of learning task characteristics. AMRL-TR 12-61-3, Aerospace Medical Research Laboratories, (January 1961), Wright-Patterson Air Force Base, Ohio.


Appendix A

Task Characteristic Rating Scales

Note: The 15 scales used for the Key West study are indicated by asterisks (*). All of the 18 indices presented in this Appendix were employed in the post-diction study.
1. NUMBER OF OUTPUT UNITS (UNIT)

The entire purpose of the task is to create output units. An output unit is the end product resulting from the task. Output units can take different forms. For example, sometimes the output unit is a physical object assembled from several parts. It may also take the form of a relationship between two or more things, e.g., drive three car-lengths behind the car in front of you. An output unit might also be a destination, e.g., run from here to the corner, with the corner being the destination.

First, identify what the output unit is in the present task. Now, count the number of such output units that someone performing this task is supposed to produce. Use the designation AMAP (As many as possible) where no actual limit exists.

2. DURATION FOR WHICH AN OUTPUT UNIT IS MAINTAINED (DURA)

Once the operator has produced an output unit, he may be required to maintain or continue it for one of several time periods. For example, it can be maintained for as long as possible. Another alternative is that completing one output unit is a signal to leave it and go on to produce the next output unit. Or, having produced the output unit, performance ends.

Choose which of the following alternatives applies here:

1) Maintain unit as long as possible.
2) Maintain unit as long as possible but continue to produce additional units.
3) Leave unit and go on to produce next unit.
4) Production of unit: signals end of task.

3. NUMBER OF ELEMENTS PER OUTPUT UNIT (ELEM)

One way of describing an output unit is in terms of the number of elements involved in its production. By elements we mean the parts or components which comprise the output unit. In an addition problem, for example, the numbers to be added are the elements which comprise the output unit. In a more physical task, the elements could be parts to be assembled or apparatus to be manipulated.

Count the number of different displays and controls which are manipulated in producing a single output unit.

71
04. SIMULTANEITY OF RESPONSES (SIMU)

The responses which the operator makes in producing one output unit may involve one or more effectors (e.g., hand, foot, arm, voice, etc.). Depending upon the task, these effectors may or may not be used simultaneously. For example, both hands (two effectors) are used simultaneously in playing a piano.

How many effectors are being used simultaneously during the present task?

zero  ______  two  ______  three  ______  four  ______

05. NUMBER OF RESPONSES (NO, R)

Earlier we were concerned about the number of elements, i.e., objects or components, involved in the production of one output unit. Now we want to consider the number of responses needed to produce one output unit. There isn't a necessary one-to-one relationship between objects and responses.

Count the number of responses or steps involved in producing one output unit for the present task. Enter this number on the answer sheet.

06. RAPIDNESS OF FEEDBACK (FEED)

For present purposes the term FEEDBACK refers to information which an operator may get about the correctness of a response. Consider the maximum number of responses the operator makes before receiving feedback on the status of output unit. Enter that number on the answer sheet.
97. WORK LOAD (LOAD)

Work load refers to the number of output units to be produced relative to the time allowed for their production. We are interested in the ratio of the number of output units per unit time, e.g., make 5 widgets in 10 minutes; 1 widget produced every two minutes.

However, there are those tasks in which the goal is to maintain a situation rather than to produce multiple output units. For example, a driving task where you are to stay within 30 feet of the vehicle ahead of you. For these types of tasks, work load refers to the length of time for which maintenance is required. The longer the maintenance period, the higher the work load.

Therefore, rating a task in terms of work load resolves to answering one of two questions:

1) How much has to be produced in what amount of time; or
2) How long does this situation have to be maintained or continued?

Definitions

High work load - as many output units as possible are to be produced in a fixed period of time; a relatively large number of output units is to be produced in a relatively short period of time; an output unit is to be maintained for a relatively long time or for as long as possible.

Examples

- Drive as many nails as possible in five minutes.
- Maintain a stimulus-control relationship as long as possible.

Moderate work load - a moderate number of output units is to be produced in a reasonable period of time; an output unit is to be maintained for a moderate period of time relative to other possible periods.

Examples

- Drive ten nails in five minutes.
- Maintain a stimulus-control relationship for three minutes.

Low work load - a small number of output units is to be produced in a relatively long period of time; an output unit is to be maintained for a relatively short period of time.

Examples

- Drive these two nails in the next five minutes.
- Sum the following five numbers.
- Maintain a stimulus-control relationship for 30 seconds.
3. PRECISION OF RESPONSES (PREC)

Tasks may differ in terms of how precise or exact the operator's responses must be. Judge the degree of precision involved in the present task by considering the most precise response made in producing an output unit.

Definitions

High degree of precision - because of small targets, fine scales, sensitive controls, etc., the subject must make responses which are extremely precise.

Moderate precision - relative to the definitions above or below, a moderate degree of precision must accompany subject's responses.

Low degree of precision - because of large targets, gross scales, insensitive controls, etc., the subject can make responses which are gross or imprecise.

Examples

- Using a chemical balance (scales) determine the weight of the following objects to the nearest microgram.
- Replace the mainspring in this wrist-watch.
- Using your pencil, trace this maze.
- Do twenty push-ups.
- Sort the oranges and lemons into two piles.
59. RESPONSE RATE (RATE)

Responses can be made at different rates. That is, the frequency with which responses must be made can vary from task to task. For example, you would have a higher rate of responding if you were playing a match of tennis than if you were playing chess. The responses would come more frequently in the first case than in the second. You are to judge what rate of responding is called for in producing one output unit in the task being judged.

<table>
<thead>
<tr>
<th>Definitions</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High rate of responding</strong> - many responses are required per unit time. In the extreme case, responses become continuous.</td>
<td>• Fire 20 rounds for effect as quickly as possible.</td>
</tr>
<tr>
<td></td>
<td>• Complete this jigsaw puzzle as fast as you can.</td>
</tr>
<tr>
<td></td>
<td>• Track this target.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Moderate rate of responding</strong> - a moderate number of responses are required per unit time.</td>
<td>• Fire 20 rounds. Fire rapidly but also be as accurate as you can.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Low rate of responding</strong> - few responses are emitted per unit time. Responses are often singular.</td>
<td>• Add the following numbers. Take all the time you need.</td>
</tr>
</tbody>
</table>

75
Consider again the number of steps (responses) involved in producing one unit. The steps may be described in terms of the dependency among them. Dependency concerns the extent to which the steps must be done in a prescribed order. For example, dependency exists between Step A and B if Step B cannot be accomplished without Step A being done first. Note: Procedures which have only one step are automatically free of dependency. Natural dependency refers to dependency that is inherent in the operation of the equipment.

Definitions

High dependency among steps: Each step in the procedure is completely dependent upon the preceding procedural step. Systematic ordering of steps is at a maximum.

Moderate dependency among steps: In the total number of steps comprising the procedure, approximately 50% are dependent upon preceding steps.

Low dependency among steps: Procedural steps are not required in any particular sequence. Step A may precede B; step B may precede A. Procedures having one step are low in dependency.

Examples

1. Using the combination you've been given, open the safe.
   • Dial this telephone number.
2. Using colored blocks, stack them into columns four blocks high. Order of color is unimportant.
   * Using colored blocks, stack them into columns four blocks high. Do this in the order red and green for the first two blocks. The remaining blocks may be of any color.
**11. TUTORIAL DEPENDENCY OF RESPONSES (TUDR)**

Consider again the number of steps (responses) involved in producing one output unit. The steps may be described in terms of the dependency among them. Dependency concerns the extent to which the steps must be done in some specified order. For example, dependency exists between steps A and B if step B cannot be accomplished without step A being done first. **Note:** Procedures which have only one step are automatically low in dependency. Tutorial dependency refers to a dependency increased as part of the training in an effort to standardize trainee operations.

<table>
<thead>
<tr>
<th>Definitions</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High dependency among steps</strong> - each step in the procedure is completely dependent upon the preceding procedural step. Systematic ordering of steps is at a maximum.</td>
<td>Using the combination you've been given, open the safe.</td>
</tr>
<tr>
<td></td>
<td>- Dial this telephone number.</td>
</tr>
<tr>
<td><strong>Moderate dependency among steps</strong> - in the total number of steps comprising the procedure, approximately 50% are dependent upon preceding steps.</td>
<td>Using colored blocks, stack them into columns four blocks high. Do this in the order red and green for the first two blocks. The remaining blocks may be of any color.</td>
</tr>
<tr>
<td></td>
<td>- Using colored blocks, stack them into columns four blocks high. Order of color is unimportant.</td>
</tr>
<tr>
<td><strong>Low dependency among steps</strong> - procedural steps are not organized in any particular sequence. Step <strong>A</strong> may precede <strong>B</strong> or <strong>B</strong> may precede <strong>A</strong>. Procedures having one step are low in dependency.</td>
<td>- Using colored blocks, stack them into columns four blocks high. Order of color is unimportant.</td>
</tr>
</tbody>
</table>
**12. OPERATOR CONTROL OF THE RESPONSE (OCOR)**

Given the occurrence of the stimulus, what degree of control does the operator have over when he must initiate his response?

<table>
<thead>
<tr>
<th>Definitions</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full operator control</strong> - the operator is the sole determiner of when the response will be made.</td>
<td>• Playing a game of chess by yourself where you play both sides and there is no time limit for responding.</td>
</tr>
<tr>
<td><strong>Partial operator control</strong> - the response must be made within a reasonable time after the stimulus occurs but the operator determines when within the interval the response will take place.</td>
<td>• The traffic light turns red when you are 500 yards from it; you have options as to when you will hit the brake.</td>
</tr>
<tr>
<td><strong>No operator control</strong> - the operator must respond as soon as the stimulus occurs.</td>
<td>• Typical reaction time task. When the light comes on, push this button as fast as you can.</td>
</tr>
</tbody>
</table>
### Definitions

<table>
<thead>
<tr>
<th>Feedback Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate feedback</td>
<td>finding the correct switch to turn on a light.</td>
</tr>
<tr>
<td>Delayed feedback - operator</td>
<td>opening a combination lock using five numbers.</td>
</tr>
<tr>
<td>No feedback provided</td>
<td>student takes a mid-term exam but is not told what grade he got.</td>
</tr>
</tbody>
</table>

**Immediate feedback** - Operator knows whether the response was correct as soon as it was completed.
14. **DEGREE OF MUSCULAR EFFORT INVOLVED (MUSC)**

This dimension considers the amount of muscular effort required to perform the task. Examine the task and identify the most physically strenuous part of it. Rate this part on the scale below.

<table>
<thead>
<tr>
<th>Definitions</th>
<th>Examples</th>
</tr>
</thead>
</table>
| High amount of muscular effort - response(s) require a high degree of muscular involvement. | • Do 40 push ups.  
• Lift the heaviest weight possible. |
| Moderate amount of muscular effort required for the response(s) | • Tighten nuts on bolts securely with a wrench. |
| Low amount of muscular effort required            | • Solder two wires together  
• Add numbers and report the sum aloud.  |
15. OPERATOR CONTROL OF THE STIMULUS DECS

What degree of control does the operator have over when the stimulus occurs or becomes relevant?

**Definitions**

- **Full operator control** - the operator is the sole determinant of when the stimulus occurs or when it becomes relevant.

- **Partial operator control** - the operator has some control over when the stimulus occurs or becomes relevant.

- **No operator control** - the operator has no control over when the stimulus occurs or when it becomes relevant.

**Examples**

- Full operator control: When bird appears.
- Partial operator control: Controlling the speed of the car to have a green light when it gets to the intersection.
- No operator control: Waiting for the telephone to ring.
16. REGULARITY OF STIMULUS OCCURRENCE (ROS0)

Consider the critical stimulus or stimulus complex to which the operator must attend. Does it occur at regular (i.e., equal) intervals or at irregular intervals. Treat regular intervals and constant presence of the stimulus as equivalent conditions.

Rate the present task on this dimension.

### Definitions

| High regularity - stimulus occurs at regular intervals or is constantly present. |
| --- | --- |
| Medium regularity - stimulus occurs at irregular (unequal) intervals but there is a pattern of occurrence. |
| Low regularity - stimulus occurs at very irregular (almost random) intervals. |

### Examples

- Cars coming along an assembly line.
- Looking at a photograph of an object.
- Receiving morse code.
- Detecting random signals on a CRT display.
Consider the critical stimulus or stimulus-complex to which the operator must attend in performing the task. Relative to the total task time, for how long a duration is the stimulus or stimulus-complex present during the task?

<table>
<thead>
<tr>
<th>Definitions</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Long duration</strong> - stimulus would remain indefinitely.</td>
<td>7 - <em>Drawing a picture by observing a model of the object being drawn.</em></td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td><strong>Medium duration</strong> - stimulus remains present until changed (spatially, temporally, etc.) by the response made to it.</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td><strong>Short duration</strong> - stimulus ceases prior to response being made to it.</td>
<td>1 - <em>Operator must identify words or targets presented to him accurately.</em></td>
</tr>
</tbody>
</table>
18. **VARIABILITY OF STIMULUS LOCATION (IVARS)**

Judge the degree to which the physical location of the stimulus or stimulus complex is predictable over task time.

<table>
<thead>
<tr>
<th>Definitions</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>High predictability - stimulus</td>
<td>• Stimulus is a red light located on a display panel.</td>
</tr>
<tr>
<td>location remains basically</td>
<td></td>
</tr>
<tr>
<td>unchanged.</td>
<td></td>
</tr>
<tr>
<td>Medium predictability - location</td>
<td>• Visually following an arrow in flight toward a target.</td>
</tr>
<tr>
<td>changes but in a known manner or</td>
<td></td>
</tr>
<tr>
<td>pattern.</td>
<td></td>
</tr>
<tr>
<td>Low predictability - location</td>
<td>• Predicting which leaf will fall off on a tree next.</td>
</tr>
<tr>
<td>changes in an almost random</td>
<td></td>
</tr>
<tr>
<td>fashion.</td>
<td></td>
</tr>
</tbody>
</table>
Appendix B

Operations Flow-Charts
Figure 2.3: Operations Flow-Chart for the 1LP Detection Sub-Task
Figure 3.7: Operations Flow-Chart for the Holo Detection Sub-Task
Figure 2.10. Operations Flow-Chart for the AN/SQS-54CX Set-Up Sub-Task.
Figure 2.11. Operations Flow-Chart for the AN/SQS-26CX Classification Sub-Task.
Appendix G
DEI Link Charts
Figure 3.1. M723 Set-Up (original) Link Chart
Figure 3.1. N/A
Figure 3.5. MS3 interaction link chart.
Figure 3.6. FISH Detection Diagram.
Figure 3.9. VARI Localization Link System
Appendix D

Prediction Study References
Prediction References


Fleishman, E. A. Unpublished data on two-hand coordinator.


Lincoln, R. S. Learning and retaining a rate of movement with the aid of kinesthetic and verbal cues. *Journal of Experimental Psychology*, 1956, 51, 199-201.


115

This study yielded two groups and hence two sets of learning data for the post-diction study.
13 ABSTRACT
An exploratory study was undertaken, as part of a program to develop quantitative techniques for prescribing the effectiveness of training systems. As a first step in this program, the present study attempted to validate an initial set of quantitative indices.  The indices were then used to describe a sample of training techniques and their effects. From this new data, a prediction methodology using these indices was developed.

The data collection included the collection of quantitative data on task performance, lay-out indices, and a set of task rating scales.  The indices were applied to task performance data, collected on seven experimental groups in a junior training school, Key West, Florida.  Application of the indices was compared with differences between groups, and differences among training devices, and within four groups of both advantages, differences in training, and limitations, classification was possible.

The predictive method which was operated was an adaptation of the standard multiple regression model.  Mean and group performance were collected on the indices, and significant multiple correlations were used in the predictive model.  A significant amount of variance in the correlational data was explained by the task performance data.  The predictive model was also shown to be highly accurate in predicting the performance of a new training technique.

II SUPPLEMENTARY NOTES
None

12 SPONSORING MILITARY ACTIVITY
Naval Training Center
Key West, Florida
**CLASSIFIED**

**Security Classification**

---

<table>
<thead>
<tr>
<th>Training Definitions</th>
<th>Training Methods</th>
<th>Grant Review and Analysis</th>
<th>GR Charactertistics</th>
<th>Performance Radiation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**RESTRICTIONS**

1. **ORIGINATING ACTIVITY.** Enter the name and address of the contractor, subcontractor, grantee, Department of Defense agency or other organization concerned with the report.

2. **REPORT SECURITY CLASSIFICATION.** Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking shall be in accordance with appropriate security regulations.

3. **GROUP.** Automatic downgrading as specified in DoD Directive 5200.15 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, when the report contains material not to be read aloud, enter the restriction as identified by security and installation directives.

4. **REPORT TITLE.** Enter the complete report title in all capital letters. Titles in all cases shall be unclassified. The meaningful title, when not classified, shall be used on all copies in parentheses immediately following the title.

5. **DEFINITIVE NOTES.** If appropriate, enter the report's content, internal progress, summary, annual, or final. Enter the incremental date, when a specific reporting period is indicated.

6. **AUTHORS.** Enter the name(s) of author(s) as shown on the report. First name, first name, middle initial, last name, street address, city, state, and zip code. The name(s) of firms supplying an associate technical report.

7. **REPORT DATE.** Enter the date of the report as day, month, year, or month-day-year. For earlier dates, enter the date as day, month, year.

8. **TOTAL NUMBER OF PAGES.** The total number of pages should follow normal pages in parentheses, i.e., enter the number of pages containing information.

9. **NUMBER OF REFERENCES.** Enter the total number of references cited in the report.

10. **CONTRACT OR GRANT NUMBER.** If appropriate, enter the applicable number of the contract or grant under which the report was written.

11. **PROJECT NUMBER.** Enter the appropriate project, program, task order, or key word number that best describes the report.

12. **CLASSIFICATION REPORT NUMBER.** Enter the official report number that is specific and shall be identified and controlled by the sponsor, as applicable. The number must include a volume number.

13. **OTHER REPORT NUMBERS.** If the report has been borrowed, copied, or distributed other than the original, indicate the number of copies distributed that are different.

14. **AVAILABILITY LIMITATION NOTES.** Enter any limitation on further dissemination of the report, either than those imposed by security classification, using standard statements such as:

   (1) "Qualified requesters may obtain copies of this report from a DDC.
   (2) "Foreign dissemination of this report by DDC is not authorized.
   (3) "U.S. government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through...
   (4) "DoD mission agencies may obtain copies of this report directly from DDC. Other qualified users shall request through...
   (5) "All distribution of this report is controlled. Qualified DDC users shall request through...

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

15. **SUPPLEMENTARY NOTES.** Use for additional explanatory notes.

16. **SPONSORING MILITARY ACTIVITY.** Enter the name of the departmental project officer or laboratory sponsoring project for the research and development, include address.

17. **ABSTRACT.** Enter a brief statement that briefly defines and describes the research, giving the scope of the research. The report should be directly cited in the body of the technical report. If not printed, include a continuation of the abstract.

For blank and other short abstracts of classified reports, an introduction of the data shall be provided, as shown in this example: 'The sponsor's report contains a brief abstract of the classified report.'

References are not to be placed in the abstract. Do not reference any pages from DDC documents.

18. **KEY WORDS.** Enter a list of meaningful terms that clearly describe the report and can be used as an index to the report. The key words should be at least four characters long each word.

19. **UNCLASSIFIED SECURITY CLEARANCE.** Enter the name of the report and the number of copies that have been distributed.

---

**UNCLASSIFIED**

**Security Classification**