DOCUMENT RESUME

ED 087 478 IR 000 230

AUTHOR	Mirabella, Angelo; Wheaton, George R.
TITLE	Effects of Task Index Variations On Transfer of Training Criteria. Final Report.
INSTITUTION	American Institutes for Research in the Behavioral Sciences, Silver Spring, Md.
SPONS AGENCY	Naval Training Equipment Center, Orlando, Fla.
REPORT NO	NAVTRAEQUIPCEN-72-C-0126-1
PUB DATE	Jan 74
NOTE	102p.; This document contains 91 leaves, some of which are tables, 11 inches wide by 8 1/2 inches high and require two microfiche frames
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	Educational Research; Military Training; Prediction; *Simulators; *Skill Development; Task Analysis; *Task Performance; Training; *Training Techniques; *Transfer of Training

ABSTRACT

The concluding series of a research program designed to validate a battery of task indexes for use in forecasting the effectiveness of training devices is described. Phase I collated 17 task indexes and applied them to sonar training devices, while in Phase II the 17 index battery was validated, using skill acquisition measures as criteria. Training of procedural skill was carried out in a modularized, synthetic sonar trainer. Significant multiple correlation coefficients were obtained for performance time and errors during skill acquisition. Phase III validated the index battery against transfer of training criteria, for the results demonstrated that quantitative variations in task designed related to variations in transfer of training measures. A set of predictive equations was constructed, and it was concluded that these equations could be used to compare trainer prototypes, although additional field validation was recommended. It was also concluded that the battery could be used in research on the interaction of task and other variables. Training method as a function of task complexity was studied, with the results indicating that the effectiveness of dynamic versus static procedural training varied with a change in task parameters. (Author/PB)

Technical Report: NAVTRAEQUIPCEN 72-C-0126-1

EFFECTS OF TASK INDEX VARIATIONS ON TRANSFER OF TRAINING CRITERIA

Angelo Mirabella
George R. Wheaton

January 1974

American Institutes for Research
Silver Spring, Maryland

Contract N61339-72-C-0126
NAVTRAEQUIPCEN Task No. 1752-03

NAVAL TRAINING EQUIPMENT CENTER
ORLANDO, FLORIDA 32813

Technical Report: NAVTRAEQUIPCEN 72-C-0126-1

EFFECTS OF TASK INDEX VARIATIONS ON TRANSFER OF TRAINING CRITERIA

ABSTRACT

The present report describes the concluding series of studies in a three-phase program of research. The overall goal of the program has been to develop and validate a battery of quantitative task indices for use in forecasting the effectiveness of training devices.

In Phase I of the program, indices were collated and applied to an assortment of passive- and active-sonar training devices. On the basis of these field applications, an initial set of 53 quantitative task indices was reduced to 17 measures.

In Phase II of the program, the 17-index battery was validated using skill acquisition measures as criteria. In this validation effort, training of procedural skill was carried out in a modularized, synthetic sonar trainer. The modular construction of the device permitted its configuration into a large number of research tasks. Substantial and significant multiple correlation coefficients were obtained for both performance time and errors during skill acquisition.

Phase III, described in the current report, extended the work of Phase II by validating the index battery against transfer of training criteria. Phase III results demonstrated that quantitative variations in task design could be related significantly and substantially to variations in transfer of training measures.

On the basis of these results and those of Phase II, a set of predictive equations was constructed.

It was concluded that these equations could be employed immediately to compare the efficacy of competing trainer prototypes, but that additional validation efforts in the field were necessary in order to extend confidence and generality of the methodology.

It was further concluded that the battery could be useful in selecting tasks for research on the interaction of task variables and other training system variables. A demonstration of this application was carried out in which training method was studied as a function of task complexity. Results of this latter study provided some support for the hypothesis that the effectiveness of dynamic versus static procedural training varied with changes in task parameters.

FOREWORD

This is the third in a series of reports the general purpose of which is to determine the feasibility of describing, in quantitative terms, tasks that are of practical importance in Navy operations. If this be possible, and if these quantitative indices can be related to the difficulty operators experience in learning the tasks and to the amount of transfer that can be carried over to performance "on the job", important implications follow about the design of training programs and the aids and devices they include.

This series of reports demonstrates the feasibility of describing tasks in quantitative terms and of relating these quantitative indices to difficulty of learning the tasks and to the amount of transfer of training to other tasks, and presents the methods for so doing.

Future work includes the validation of the computation of the quantitative indices and of the methods for their use in an actual Navy training/operational environment. Plans are being laid to perform these validations.

The first two reports in this series are: NAVTRADEVCEN 69-C-0278-1, Trainee and Instructor Task Quantification: Development of Quantitative Indices and A Predictive Methodology, and NAVTRAEQUIPCEN 71-C-0059-1, Effects Of Task Index Variations On Training Effectiveness Criteria.

Vincent J. Sharkey
VINCENT J. SHARKEY
Scientific Officer

ACKNOWLEDGMENTS

The authors wish to thank their project monitor, Mr. Vince Sharkey, for his assistance at various stages of this research. In addition, appreciation is extended to Ms. Gloria Greenbaum, Ms. Genie Brahlek and Ms. Ellen Schaffer of AIR for their competent and enthusiastic assistance in collecting and processing the data.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd)

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (Cont'd)

SECTION I

INTRODUCTION

A number of complex problems confront individuals who are responsible for the design and development of effective training devices. One of the most difficult to resolve is the problem of task fidelity. Early during conceptualization of the device, decisions must be made concerning those features of the operational task which should be incorporated into the trainer in order to make the device optimally effective for both the acquisition and transfer of skills. Complementary decisions are needed concerning those features of the operational task which can be cost-effectively eliminated. Yet, objective means for deciding on a priori grounds what to include and what to eliminate have never been developed. In particular, quantitative methods have been lacking with which to relate variations in trainer task characteristics to variations in the acquisition and transfer of skill. The pragmatic consequence of this situation has been incorporation into training devices--and, in particular, simulators--of as much realism as the state-of-the-art and available dollars will permit. Increasingly, the cost-effectiveness of such a response to training needs has been questioned.

A major stumbling block to the development of more objective and systematic approaches to device design has been the lack of an acceptable method for quantitatively analyzing and describing trainee tasks. In turn, two issues underlie development of the required methodology. First, is it possible to describe the critical features of a device reliably and along a number of quantitative dimensions? Unless such description is possible there will be no way to investigate the relationship of interest. Second, can measures of training effectiveness (i.e., rate of skill acquisition, level of transfer) be demonstrated to vary in some predictable manner as features of a training device are manipulated? Unless there is a relationship between these two sets of variables, prediction of effectiveness will not be feasible.

BACKGROUND

To resolve these issues the Naval Training Equipment Center (NAVTRA-EQUIPCEN) sponsored the American Institutes for Research in a program of research which was executed in a series of phases. The goals of the program were to: (1) develop or compile a set of quantitative task descriptive indices; (2) determine the feasibility of using such indices to describe different kinds of trainee tasks; and (3) explore the relationship between such indices and measures of skill acquisition and transfer of training. The phases of research conducted in support of these goals are summarized below.

PHASE I - DEVELOPMENT OF QUANTITATIVE INDICES. The first phase of the research program had three objectives. The first was to compile an initial set of quantitative indices relating to selected characteristics of various man-machine tasks. The second was to determine whether the obtained indices could be used to describe a sample of trainee tasks and to differentiate among them. The third was to develop a predictive methodology based upon the task indices and to assess its potential utility.

To accomplish these ends, the first step taken was to review the spectrum of Navy training devices 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 they contained trainee and instructor tasks which were reasonably formalized and invariant with respect to the equipment and procedures used. 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.). The decision was made, therefore, to focus initially on a more homogeneous subset of devices. This approach was adopted because it was felt that focus on a specific subset 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. On this basis Navy sensor-based or surveillance systems were chosen for study, including such devices as sonar, radar, and electronic countermeasures trainers. While attention was focused specifically on sonar trainers, the intention was to generate indices which would also provide for the quantitative description of other devices within the surveillance family.

The next step was to analyze the trainee tasks associated with these devices in detail, in order to determine the major sub-tasks performed by trainees, and to obtain information about those features of the sub-tasks which might provide a basis for generation of descriptive indices. Evaluation of several devices resulted in identification of four major trainee sub-tasks 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.

In selecting and developing quantitative indices to be used in describing the four trainee sub-tasks, consideration was given to critical task characteristics which, if manipulated, could be hypothesized to exert an appreciable effect upon rate of acquisition or level of proficiency. Based upon an examination of the four sub-tasks and upon a review of the literature, 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) a set of task characteristic rating scales; (2) the Display Evaluative Index; and (3) a set of panel lay-out and task-type indices. The second set contained specific indices which were developed to provide for 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.

As described in the Phase I report (Wheaton, Mirabella, and Farina, 1971) the 13 task characteristic rating scales were selected from a larger set of 19 scales originally developed during the course of an AIR taxonomy project (Fleishman, Teichner, and Stephenson, 1970). The scales were specifically designed to describe tasks per se, independent of two 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, responses, procedures, stimuli and stimulus-response relationships. Several rating scales were developed for each of these components. A complete discussion of the task characteristic approach is given in a report by Farina and Wheaton (1971).

The Display Evaluative Index (DEI) is a measure of the effectiveness with which information flows from displays via the operator to corresponding controls. The index, developed by Siegel, Miehle, & Federman (1962a), 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. 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, et al., 1962a; Siegel & Federman, 1967). Moreover, the index has been partially validated, i.e., against judgments by human engineering experts (Siegel, et al., 1962a; 1963).

The panel lay-out indices of Fowler, Williams, Fowler, & Young (1968) are designed to provide description of two different aspects of a man-machine task. One set is used to measure 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 to which different operations or "task types" are embodied in a particular operator console. These indices can vary independently of the DEI, which does not address itself to panel arrangements or types of panel operations. During Phase I eight of these types of indices were investigated.

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 (Folley, 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 & Harabedian, 1964). Two additional indices were the number of redundant information sources processed simultaneously (Mirabella, 1969), 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.

In addition to the generic indices, which cut across both training devices and trainee sub-tasks, an additional set of 25 descriptors was used. Fifteen of the 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. They included such items as signal persistency and display-control ratios. An additional set of ten descriptors related to the use of different training techniques. These included statements, for example, about the use of training tapes, adaptive techniques, part-task training, problem freeze techniques, etc. Altogether, 29 generic indices, 15 specific indices, and ten training technique descriptors were assembled.

The indices were applied to detailed task-analytic data collected on three sonar devices, each of which incorporated the four basic sub-tasks. In general, application of the DEI was straightforward. Values could be obtained fairly quickly, reliability did not appear to be a problem, and the index differentiated sub-tasks and devices. The panel lay-out indices also differentiated between and within sub-tasks, although they appeared to be rather labile. Several were difficult to apply and their reliability was questionable. Other generic indices, including several of the rating scales, did not appear to provide for adequate differentiation among devices. Overall, though, results were encouraging with respect to the generic indices.

The results from applying the 15 specific and ten training technique indices were generally inconclusive. Many specific indices could not be applied; when they could be, they did not clearly discriminate among tasks or devices. Training indices were simply binary statements about the presence or absence of a "freeze" capability, for instance.

In conclusion, Phase I research demonstrated the feasibility of using a variety of quantitative indices to describe salient characteristics of actual trainee sub-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 were directly related to features of a task familiar to design engineers. These were 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 could be hypothesized to bear a relationship to measures of task performance including rates of skill acquisition.

In theory, therefore, the possibility existed of developing quantitative profiles of tasks and of relating such profiles to measures of performance. Were information of this type available, it might 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 was contingent, of course, upon the demonstration of a relationship between the quantitative indices and measures of performance. Phase II of the program was concerned with this issue.

PHASE II - PREDICTION OF SKILL ACQUISITION. Phase II also had three objectives. The first was to refine the set of quantitative indices employed during the earlier research, adding new descriptors, if possible, while deleting those which had proved unsatisfactory. The second was to conduct an investigation of the relationship between variations in quantitative indices and corresponding changes, if any, in selected criterion measures. This effort was to be conducted in a laboratory setting in order to exercise control over other variables not of immediate interest to the present study. The third and final objective was to determine whether support for relationships established in the laboratory could be provided by data collected in the field. Such support would increase confidence in the validity of the basic methodology--that of using quantitative task index information to forecast the relative effectiveness of competing designs.

To accomplish these objectives, an approach was adopted consisting of three distinct but interrelated activities. Quantification of devices in the field was continued using a revised set of indices. The data obtained during this exercise were then used in conducting a two-pronged validation study consisting of a laboratory and a field effort.

Before either validation effort could be initiated, quantitative task index data were required on a sample of actual devices. These data were intended to provide guidelines for the types and ranges of design characteristics to be manipulated in the laboratory. In addition, they were to be employed directly in the anticipated field validation effort as the predictor variables. Accordingly, efforts begun during Phase I to apply the quantitative indices were continued. Application of the indices was extended to several devices not examined during the earlier work. Altogether, 13 different trainee stations were quantified including: the 14E10/3 at Quonset Point, Rhode Island; the 14B31B (AQA-1 and ASA-20 stations), 14E14, and X14A2 at Norfolk, Virginia; the 21A39/2 (OA1283, BQR-2C, and BQR-7 stations) at Charleston, South Carolina; and the 14E3, 14A2/C1, SQS-26CX, and 21B55 (OA1283 and BQR-2B stations) at Key West, Florida.

The trainee tasks within each of the devices were analyzed in terms of a reduced set of the total number of quantitative indices compiled during Phase I. Exclusion of indices from the reduced set occurred for one of four reasons. Some, most notably a set of task characteristic rating scales, were excluded because: (1) they were often difficult to apply reliably, requiring a consensus among several analysts; and (2) they referred in many instances to characteristics which, although varying across very different types of devices, did not appear to reflect readily manipulable design features (e.g., a work load dimension). Still other indices were excluded either because they generated little variation for the present types of devices or because they had been found from past work to be correlated highly with other descriptors. The set of descriptors finally adopted included 17 indices. These were defined in the Phase II report (Wheaton, and Mirabella, 1972).

Values were obtained on all 17 indices for each of the major trainee sub-tasks within each of the 13 devices. The index data for all four sub-tasks were used as predictors in the field validation effort. The index data obtained for the various set-up sub-tasks provided guidelines for the laboratory research.

The general approach to laboratory validation was to develop a modularized, synthetic sonar trainer, capable of being readily configured into a large number of sonar "trainers", varying in design characteristics, but with a common set of functions. The trainer was designed to evaluate set-up behavior alone. An attempt was made to compile a set of configurations which would vary as much as possible along the 17 design indices selected for study. Toward this end, three anchor configurations were chosen. There was a "complex" trainer consisting of all complex panels, a "simple" trainer consisting of all the simple panels which were available, and a medium configuration which was generated by randomly selecting either a complex or a simple module for each function on the trainer console.

In addition to these three primary trainers, nine additional trainers were selected to yield a range of design parameter values. These configurations essentially represented variations in the simple trainer or the medium trainer; i.e., the simple trainer embedded in the complex, medium trainer with feedback lights removed, simple trainer with additional contingency responses included in the training regimen. These manipulations were aimed at reducing correlations among the design parameters, in particular the correlation between number of displays or controls and other design characteristics. For each trainer, a specific set of procedures or sequence of responses was developed. These served to define "trainee" tasks analogous to the trainee set-up subtasks associated with actual sonar training devices.

Following development of the synthetic trainer and selection of the specific tasks to be studied, the testing portion of the laboratory effort was initiated. Subjects were recruited from local universities and were randomly assigned in groups of five to each of the 12 experimental tasks. The 60 subjects employed in this manner were paid for their services. Following procedures outlined elsewhere (Wheaton and Mirabella, 1972), data were collected representing subjects' time and error performance during skill acquisition. On a few tasks pilot transfer data was also obtained.

The second prong of the dual validation attempt involved a study of the effectiveness of the 13 sonar training devices which had been previously task analyzed. The field validation was pursued via structured interviews with experienced sonar instructors. These instructors were asked to rate the tasks trained on their devices against a set of "synthesized" comparison tasks. With respect to the sub-tasks found in each device, four specific judgments were to be made including: (1) training time; (2) proficiency level; (3) degree of transfer of training; and (4) level of task difficulty.

In general, the results of the laboratory validation effort were very encouraging. Significant multiple correlations were obtained between the quantitative task indices and speed and accuracy of performance during skill acquisition. Very tentative relationships were also established between some of the indices and measures of transfer of training. Support for these findings was obtained from the field validation study. Here again, significant relationships were established between instructors' judgments of training criteria and trainee task index values. It was to increase the stability of and to expand upon these predictive relationships that the present phase of research, Phase III, was undertaken.

PHASE III - RESEARCH OBJECTIVES. The third phase of the program consisted of three research objectives. Having demonstrated that quantitative task indices could be related to the acquisition of procedural task skill, refinement of the predictive relationships was in order. Accordingly, the first objective was to repeat the skill acquisition analyses using a modified set of predictors and a larger number of trainee tasks in the laboratory context. The second objective was to develop similar predictive relationships between task indices and measures of transfer of training. The possibility of such relationships was suggested by the findings stemming from Phase II research. The third and final objective was to demonstrate the manner in which a task quantification schema might be used when conducting training system research. Toward this end, a laboratory study was undertaken to examine the interaction between task complexity and method of training.

SECTION II

METHODOLOGY

The general approach used in the current phase of this research program (Phase III) was an extension of the method used in Phase II (Wheaton and Mirabella, 1972). In Phase III emphasis was placed on measuring transfer; manipulation of training regimens was also added.

As in Phase II, the experimental task was based upon a modularized synthetic sonar trainer, constructed to represent a cross section of some 13 different sonar devices which had been previously task analyzed. The trainer consisted of 20 different modular panels representing different sonar console functions. For most of the functions there were alternatively designed panels which could be interchanged, and, thus, used to manipulate the overall appearance of the trainer console. Figure 1 shows a photograph of one such console configuration. This was defined as our most complex configuration. Note, for example, the panel at the top left. This panel represents the function of energizing the console. It consists of a number of toggle switches, feedback lights, a rotary switch, and a meter. In other configurations of the console, this particular panel might be replaced by one which consists of nothing more than one toggle switch and one feedback light. Similarly, most of the other panels were designed in alternative forms: a "simple" version and a "complex" version for accomplishing basically the same function.

Through appropriate use of panels, there were a number of ways in which the operator's task could be manipulated. For instance: (1) alternative panels could be employed; (2) the trainee's task could be embedded in a more complex console configuration by making some of the displays and controls contained in the console irrelevant for performance of the task; (3) feedback lights associated with toggle switches could be masked; and (4) contingency responses could be built into the training procedure. These various manipulations were employed and then the task characteristic index battery (Appendix A) was used to describe quantitatively the resultant configurations. Twenty different tasks were generated in this manner, for each of which there was a corresponding set of task index values.

For any task, trainees were required to learn a set-up procedure. The general method of instruction was to describe to them the entire procedure, twice in succession. Each response in the procedure was indicated to the trainee, along with a verbal statement which he was to make as he performed a particular operation. For example, he was told to set the power switch, No. 1, to the "on" position, and say, "No. 1 to on". Verbalization by the trainee was necessary to facilitate the recording of incorrect or omitted responses in the subsequent test trials. The experimenter could identify these errors by following a procedural checklist, and noting where the trainee deviated from expected verbal statements. A stopwatch record of total performance time for each test trial was maintained.

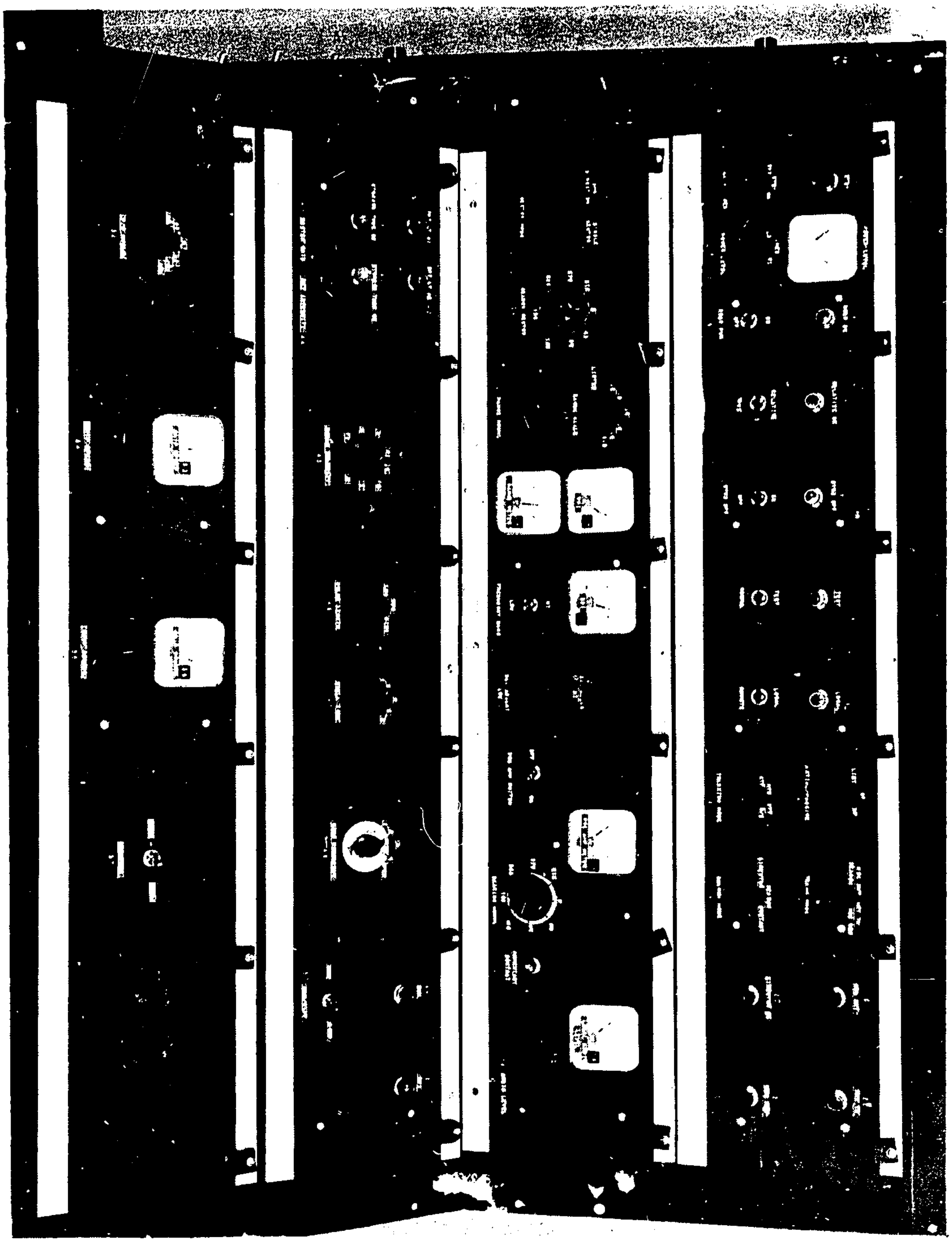


Figure 1. Complex-all Console

Following the initial two orientation trials, the trainee was exposed to 15 test trials, each involving a complete run-through of the set-up procedure for that particular task. He was interrupted for any wrong or omitted responses, and the stopwatch was halted while corrective instructions were given. It should be emphasized that following each trial the settings of all controls were scrambled so that the initial appearance of the console varied somewhat from trial to trial. Furthermore, there were a number of response sequences which could change from trial to trial as a function of experimenter inputs. As an example, the trainee might have been instructed to set up for passive-sonar search on one trial, and for active-sonar search on a subsequent trial. The specific sequence of required responses varied accordingly. Consequently, the 20 experimental tasks which were employed consisted of more than merely rote activities.

All subjects, upon completion of the initial 15 acquisition trials, transferred to a common task of medium complexity. They received one orientation trial and ten test trials on the second or transfer task. Thus, some groups of subjects transferred from difficult tasks to the intermediate task, while others transferred from relatively easy tasks to the intermediate task. Comparisons of transfer of training were based upon performance on the common intermediate task. The criteria of interest were the actual time and error scores achieved on the second or transfer task.

Each experimental group was composed of five trainees, drawn from universities in the Washington, D. C. area. Each trainee was assigned arbitrarily to only one experimental group.

STUDY 1: TRANSFER OF TRAINING

The general goal of Phase II (Wheaton and Mirabella, 1972) was to validate the 17-index battery (Appendix A), using skill acquisition as the criterion. Having succeeded in doing so, attention turned next to the issue of transfer of training. Could those same indices predict transfer and how would the specific patterns of predictors compare with those found in Phase II for acquisition? The purpose of Study 1 was to address these questions. An incidental purpose was to collect additional acquisition data in order to expand the sample used for the Phase II laboratory predictions.

PROCEDURE. For this study, twenty tasks (defined in Appendix B) were employed. However, data for nine of those tasks were carried over from Phase II. Of the nine tasks from Phase II, four included both transfer and acquisition scores. The remaining five included only acquisition scores. Thus, data were available for 15 tasks for transfer analysis and 20 tasks for acquisition analysis. Tasks were chosen with a view toward generating a wide range of task index values. At the same time, however, they were chosen to permit a preliminary study of the interactions of several of the underlying task dimensions which had been manipulated in order to generate the task index values. It was felt that such preliminary study would assist both in conducting and interpreting the regression analysis which was the focus of this investigation.

Each trainee was put through the following regimen: two preliminary training trials, followed by 15 acquisition trials, a half-hour break, and then orientation and transfer to task Ma, medium-all. Time and error measures were collected on the 15 acquisition trials and on the 10 transfer trials.

STUDY 2: INTERACTION BETWEEN TASK CHARACTERISTICS AND TRAINING METHODS

The main thrust of the program which is being concluded with this report has been upon trainee task variables. It is recognized, however, that training device utilization, and individual difference variables must, in the final analysis, all be factored into the "effectiveness" equation. Of particular potential importance are interactions among these classes of variables.

Study 2 was intended to extend our research beyond the task variable area and to demonstrate the value of looking at interactions between tasks and other variables. We chose to manipulate mode of console presentation during training since past research has indicated that dynamic presentations are not necessary for the training of procedural tasks (Grimsley, 1969; Prophet & Boyd, 1970; and Bernstein & Gonzalez, 1971). It was hypothesized that this conclusion would be dependent upon level of task complexity. More specifically, it was anticipated that dynamic presentation would be increasingly advantageous as task complexity increased.

The procedures employed were basically those of Study 1 except that the synthetic trainer was represented in one of three different ways during acquisition training.

1. "Hot" Panel. This was the dynamic mode employed in all previous laboratory work. Trainees operated the actual controls and read corresponding display values.

2. "Cold" Panel. Trainees assigned to this presentation mode operated the actual controls but were told what the display values were. All displays were inoperative.

3. Pictorial Presentation. Trainees under this condition learned their procedural task with the aid of an 11 x 14-inch photograph of the sonar trainer. They indicated control actions by pointing to appropriate positions on the photograph. Again display values were provided by the experimenter.

All subjects were then given a transfer test (10 trials) on the "hot" panel version of task Ma. Six of the twenty original synthetic sonar tasks were chosen for training in Study 2, with five trainees assigned to each combination of task and training method. Tasks included were Sa, Ma, Ca, and their embedded versions, SEma, SEca, MEca (Appendix B). This set permitted a number of different contrasts involving task complexity, task embeddedness, and training method. The organization of experimental conditions for Study 2 is shown in Table 1.

TABLE 1. EXPERIMENTAL CONDITIONS FOR STUDY 2:
TASK CHARACTERISTICS VS. TRAINING METHODS

Tasks	Training Methods		
	Hot Panel	Cold Panel	Pictorial
Ca			
Ma			
Sa			
MEca			
SEca			
SEma			

SECTION III

RESULTS

Results from both the transfer of training (Study 1) and the training method (Study 2) studies are presented in this section. The first set of analyses deals with acquisition data obtained from the synthetic set-up trainer during the course of the transfer of training study. Included within this set are analyses of variance focusing on the reliability of the acquisition data and on the interactive effects of task complexity, feedback (i.e., indicator lights), and embedding parameters on skill acquisition. The set concludes with multiple regression analyses relating task indices to acquisition time and error criteria.

The second set of analyses is analogous to the first, except that the data are transfer-of-training measures. Analyses are presented with respect to the reliability of transfer data, the interactive effects of task parameters on transfer, and the multiple regression between task indices and transfer criteria.

The final set of analyses focuses on both acquisition and transfer data from Study 2. Analyses of variance are presented which examine the interactive effects of training methods and task parameters on skill acquisition and transfer.

STUDY 1: TRANSFER OF TRAINING

Results of the acquisition and transfer portions of the transfer of training study are presented in figures 2-14 and tables 1-5. In describing both portions of this study the same format is followed. Evidence for the reliability of the data collection procedure is provided first. Second, analyses are then presented which assess the extent to which a linear regression model can be used in relating task indices to acquisition or transfer criteria. Finally, several regression analyses are then presented, some of which utilize observed interactions in the prediction equation, and some of which do not.

ACQUISITION. A number of task conditions employed in Phase II research were replicated during Phase III. Comparison of the acquisition data resulting on these two different occasions permitted some assessment of the reliability of the measures being employed. The acquisition data are shown in figures 2 and 3 for the complex-all task (Ca), the simple-all task (Sa), and the simple-all task embedded in the complex console (SEca).

Figure 2 shows mean time per trial as a function of trial block. The overlap of results for like tasks, sampled on the two different occasions, is clear. Corresponding levels of performance were obtained, in spite of the fact that different experimenters and different groups of subjects were involved.

Figure 3 shows mean number of errors in the trainee's action or verbal response as a function of trial block. In this case the overlap within each of the three tasks is still evident, although less clear-cut than for the time data shown in figure 2. Some fairly wide disparities

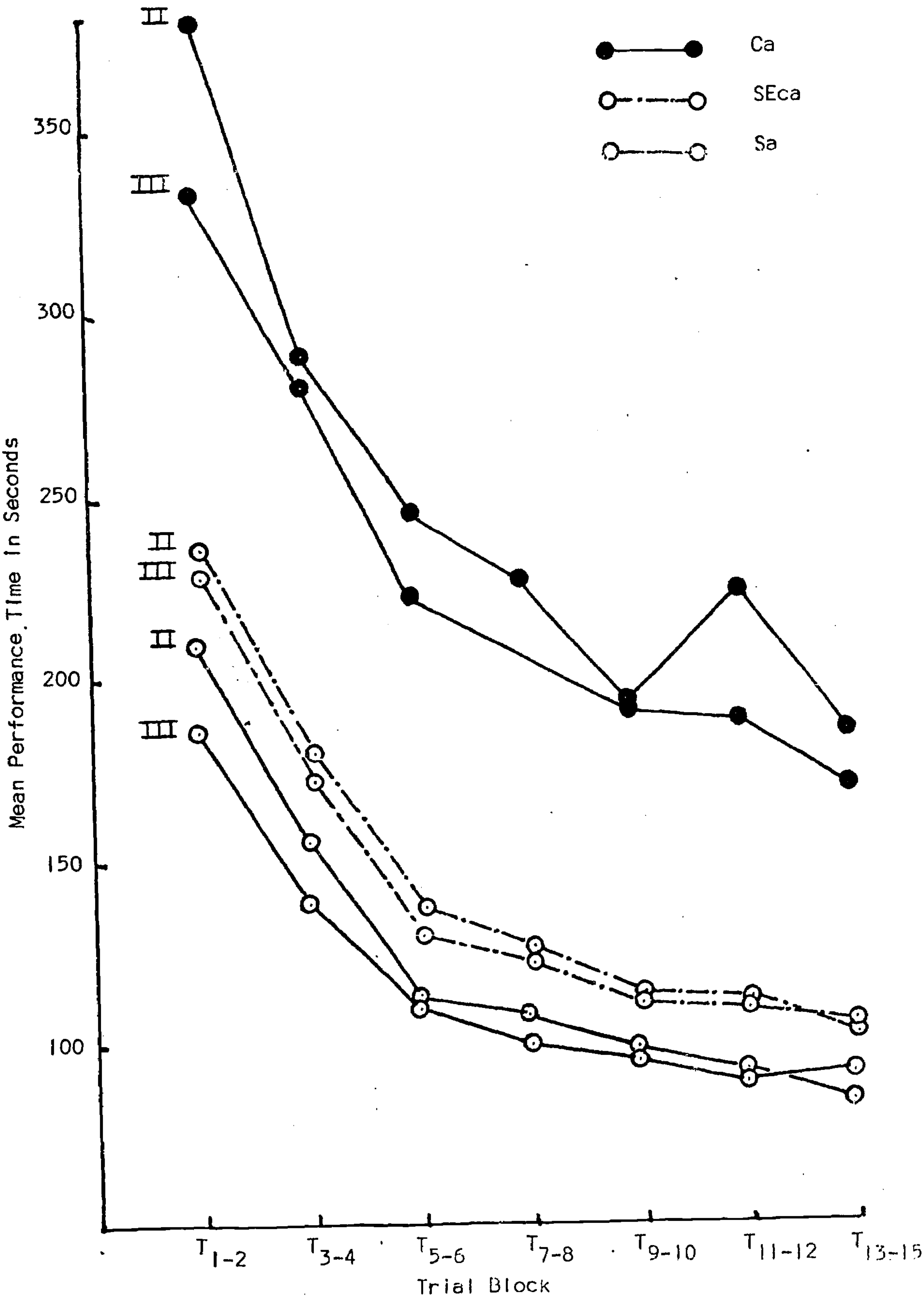


Figure 2. Mean time per trial as a function of trial block for acquisition training (Phase II and Phase III data compared for simple and complex configurations)

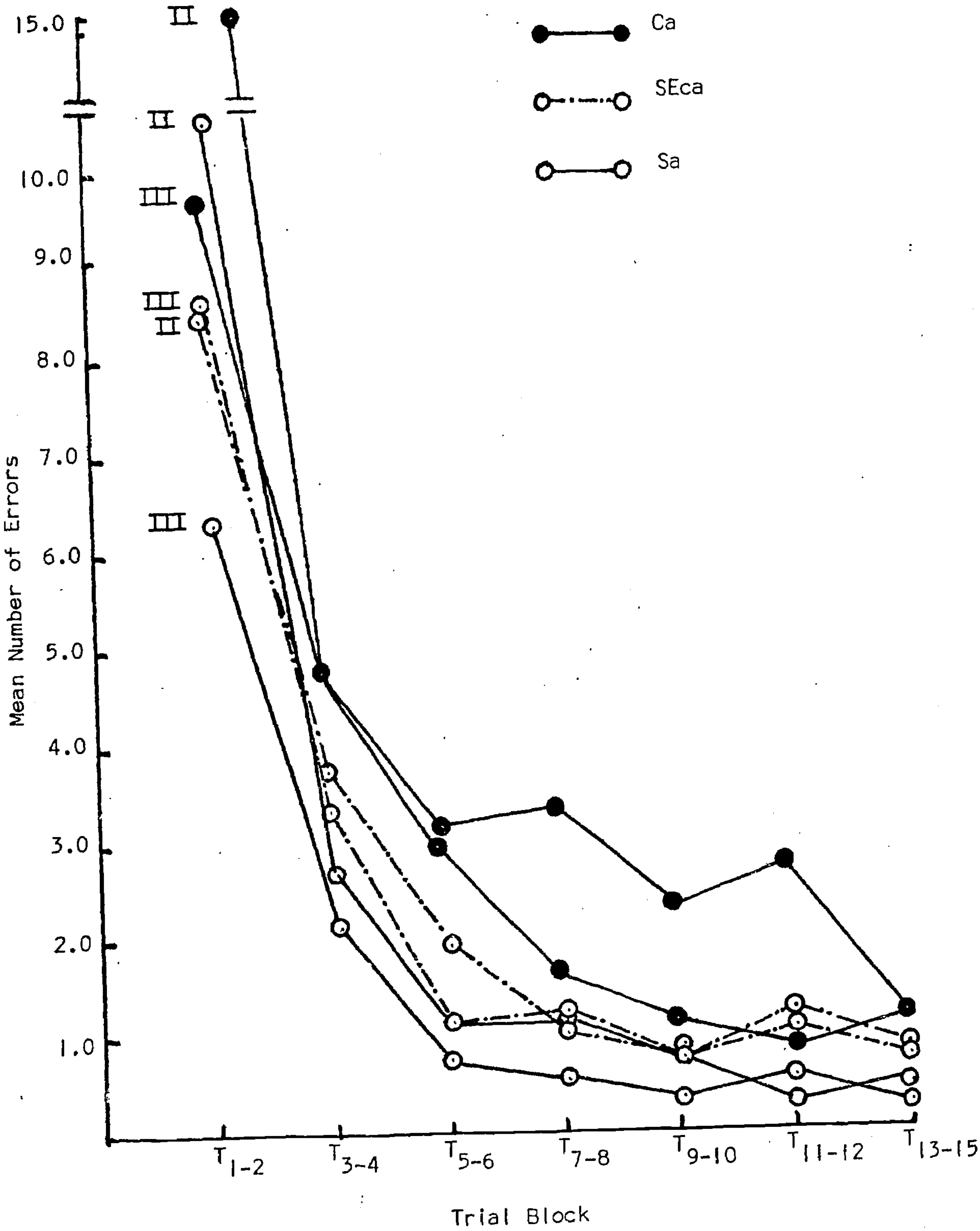


Figure 3. Mean number of errors as a function of trial block during acquisition training (Phase II and Phase III data compared for simple and complex configurations)

can be seen during the initial block of acquisition trials (i.e., T_{1-2}), but these narrow substantially for subsequent blocks. An analysis of variance conducted on the error data revealed that the overall replication effect (i.e., Phase II vs. Phase III) was not significant ($F = 3.04$; $df = 1,24$; $p > .05$).

In summary, the similarity between comparable tasks appears to be greater for the time than for the error criterion. Generally, however, both acquisition measures appear to be reasonably reliable.

Acquisition time and error measures were available for a sample of 20 different tasks, nine of these tasks having been selected from among those studied during earlier Phase II research. However, prior to use of this sample of tasks in a multiple regression analysis, subsets were selected for detailed study in a series of linear contrasts designed to highlight interactions among task parameters. Contrasts were employed which emphasized, for instance, the possible interaction between task complexity (complex, medium, simple) and amount of performance feedback (all or none); the interaction between amount of task embeddedness and degree of feedback for a fixed level of task complexity; and, combinations among all three major variables - feedback, task complexity, and embeddedness.

In a series of linear contrasts, the main effects of complexity, feedback, embeddedness, and trials were all found to influence acquisition performance, as expected. The important interactions which might influence the multiple regression model were then examined. The salient findings stemming from these analyses are represented in figures 4-9 for acquisition time and error data. Figure 4 shows mean number of errors as a function of task complexity, feedback, and trial block. There is a significant interaction between task complexity and trial block ($F = 3.86$; $df = 12,144$; $p < .001$), which can be clearly seen within either level of feedback. The initial differences in error rate associated with the various levels of task complexity, although maintained across trials, decrease as training continues. Although covariance analysis was not performed, the spread in scores appears to be substantially greater than expected on the basis of total number of task responses alone. For example, total Cn errors exceed Sn errors by 255%, but total Cn task responses exceed those for Sn by only 81%. Total Mn errors exceed Sn errors by 194%, but total Mn response actions exceed those for Sn by only 28%. Similar differences hold for the other relevant pairings.

Figure 4 also suggests a feedback by task complexity interaction. Of particular interest is the reversal in performance where feedback is removed; i.e., a greater average number of errors results from removal of feedback, even though fewer responses are required in such tasks. This mean reversal effect is greatest for the complex configuration, somewhat less for the medium configuration, and not present for the simple configuration. Statistically, however, support for an interaction between these two parameters was not obtained ($F = 1.06$; $df = 2,24$; $p > .05$).

The data for mean acquisition time shown in figure 5 generally reflect the number of responses required by the task. For example, mean

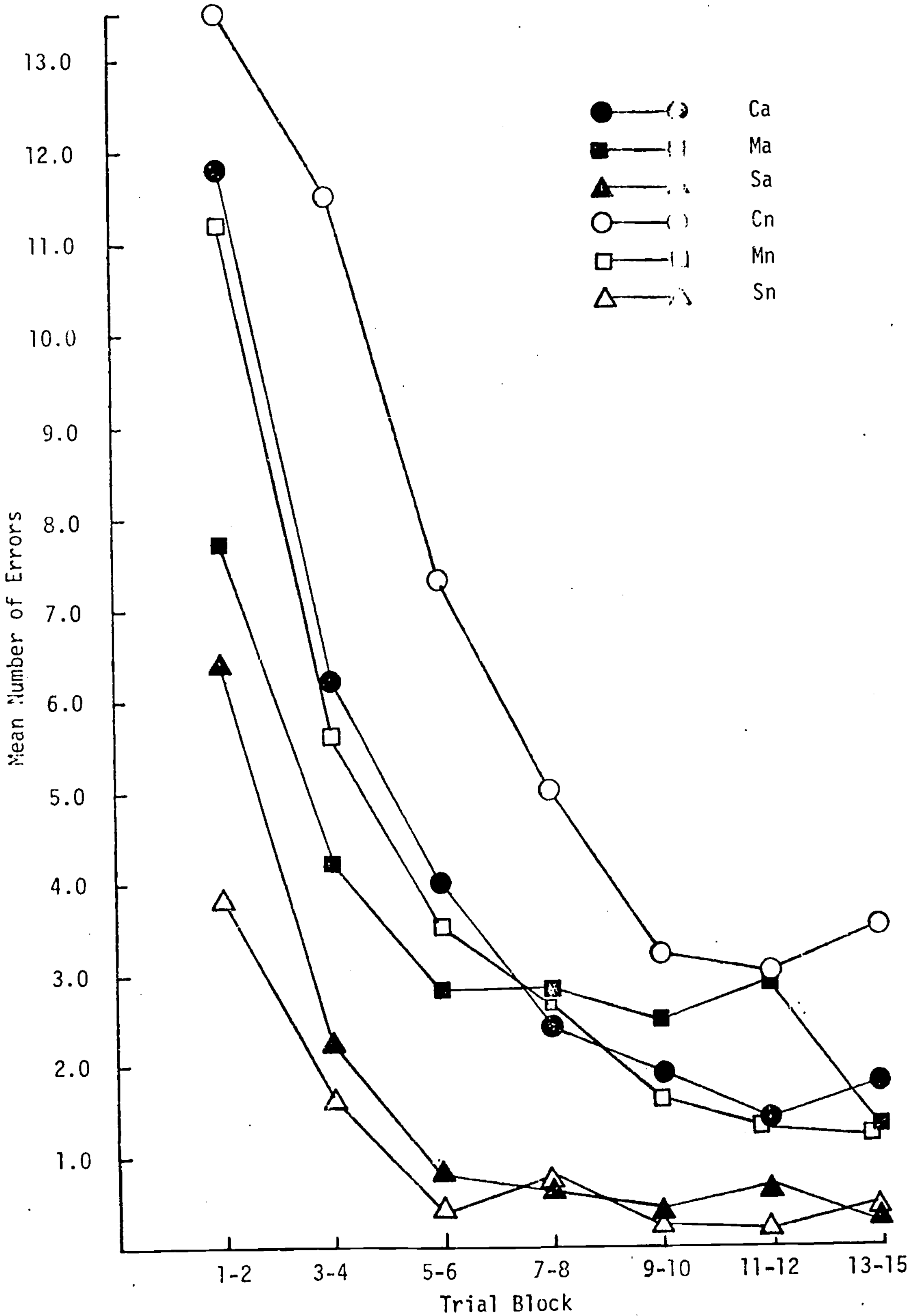


Figure 4. Acquisition errors as a function of task complexity, amount of feedback, and trial block

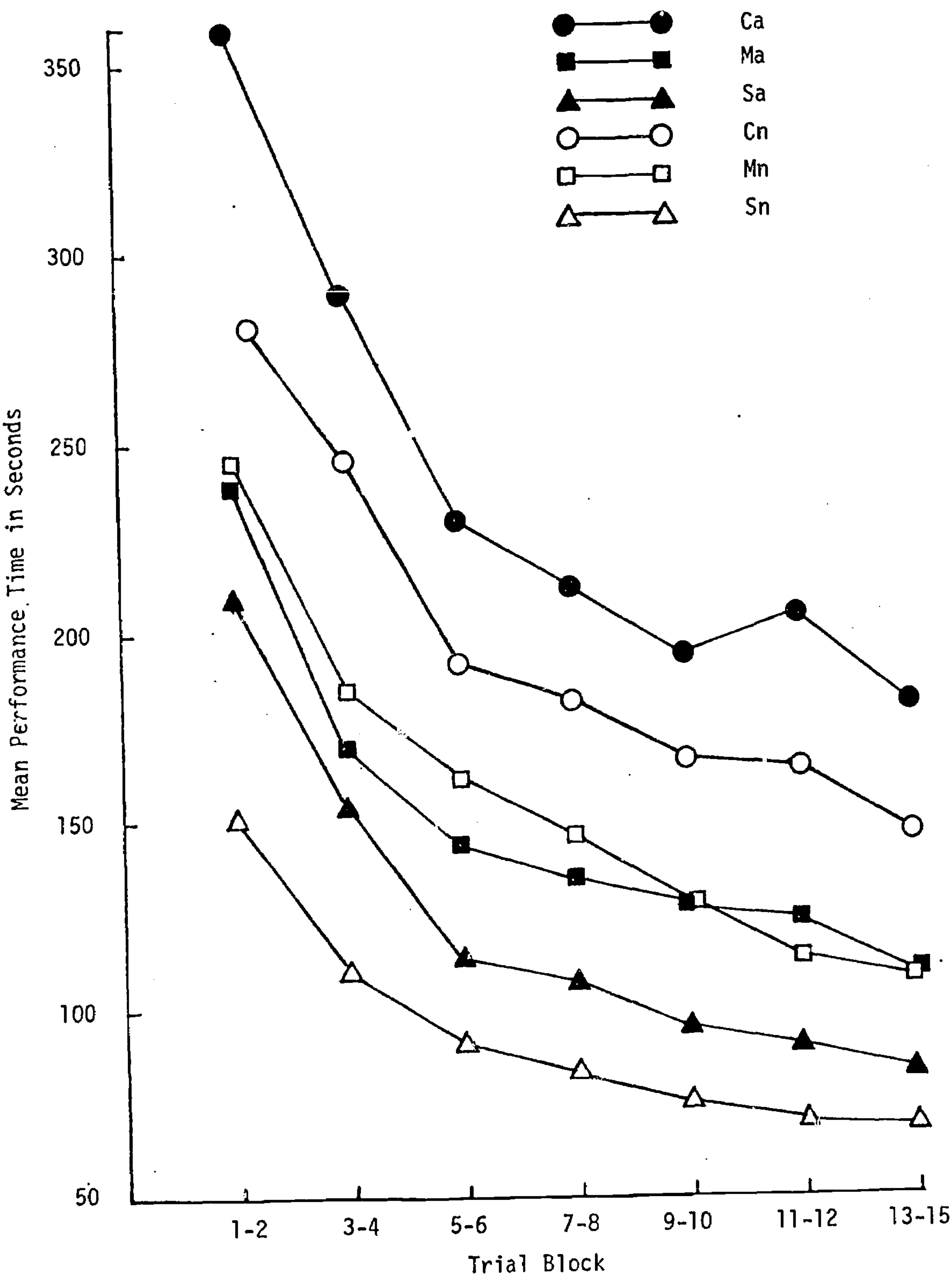


Figure 5. Acquisition performance time as a function of task complexity, amount of feedback, and trial block

Ca time is greater than mean Cn time although fewer errors (figure 4) are made on the Ca task. The initial differences in performance time due to level of task complexity decrease over training as indicated by a significant complexity by trial block interaction ($F = 2.28$; $df = 12,144$; $p < .01$). There was no indication of an interaction between complexity and feedback task parameters ($F = 0.74$; $df = 2,24$; $p > .05$).

Overall, the effects of task complexity on skill acquisition criteria are reasonably clear-cut and systematic. The more complex the task becomes, the more errors are made and the longer are performance times. Degradation in the accuracy and speed of performance increases disproportionately with increasing task responses, a finding which emphasizes the underlying multivariate nature of task difficulty or complexity.

The effects of different levels of the second major task variable, namely feedback, are presented in figures 6 and 7 for acquisition error and time, respectively. It will be recalled that, as used in this study, feedback refers to the use of certain indicator bulbs during performance of the task, a manipulation not to be confused with "feedback as knowledge of results". A significant interaction ($F = 2.06$; $df = 24,216$; $p < .005$) exists between feedback, level of embedding, and trial block for acquisition error scores as shown in figure 6. Within each level of embedding, the initial distinctions among levels of feedback decrease over trial blocks; by the end of the acquisition session all three feedback conditions exhibit essentially the same error rate. More interesting, however, is the interplay between level of feedback and degree of embedding. When the simple task is embedded in the complex console (i.e., high embedding) there is a rather consistent ordering of feedback levels. Most errors are associated with the use of all indicator lights, fewer with the use of an intermediate number of lights, and least when no indicator lights are used during task performance. When the same task is performed on a console which is fully utilized (i.e., when there is no embedding) the order is changed substantially. Most errors occur under the no-feedback condition and fewer under the high-feedback condition. Both of these levels of feedback lead to higher errors under moderate embedding than does the intermediate feedback condition.

Tentatively, at least for the procedural task used in this experiment, as the level of embedding increases, errors become a function of increasing levels of feedback. Apparently, the distinction between the task (figure) and console (background) becomes less obvious as more and more feedback indicators are used during task performance. Conversely, as the percentage of distracting stimuli decreases (i.e., there is less embedding), increasing errors are associated with decreasing feedback.

As shown in figure 7, feedback has a simpler and more systematic effect on performance time. A significant feedback by trial interaction ($F = 2.50$; $df = 12,216$; $p < .005$) exists in which initial differences due to level of feedback diminish over time. The results simply suggest that tasks consisting of more responses (e.g., high feedback in which all indicator lights are responded to) take relatively longer to perform than tasks consisting of fewer responses (e.g., tasks in which indicator lights are eliminated).

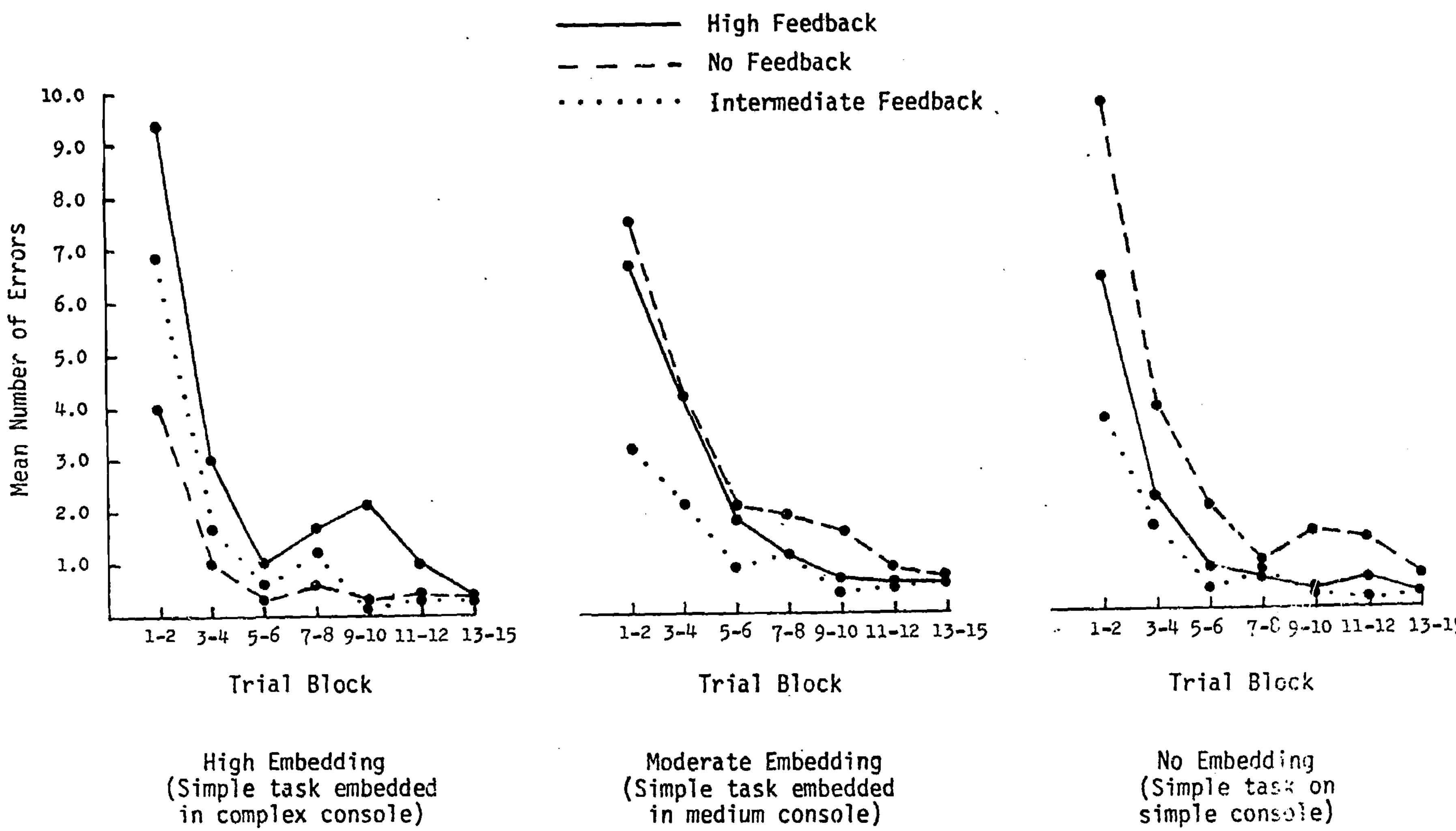


Figure 6. Mean acquisition error as a function of feedback, embedding, and trial block

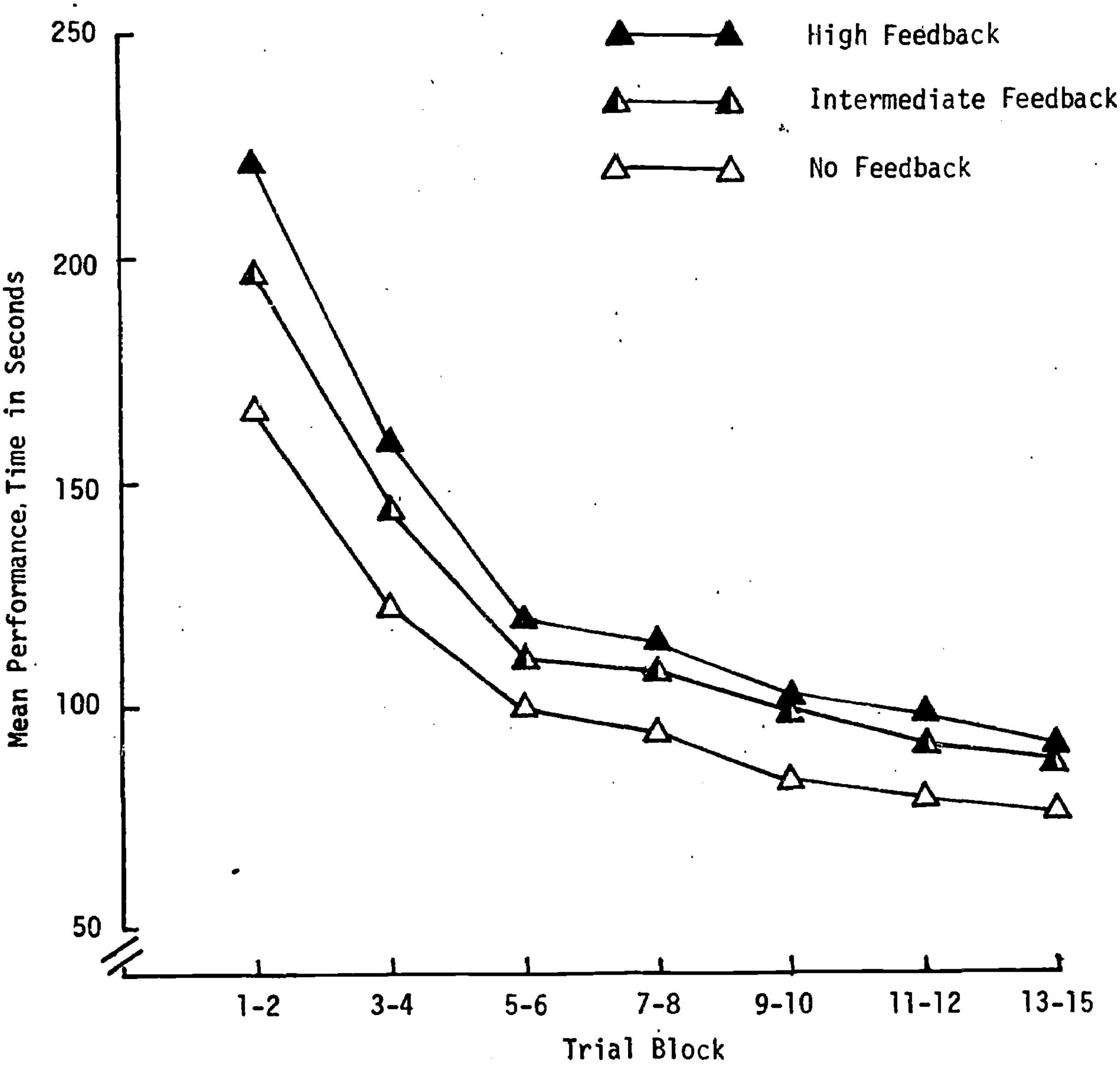


Figure 7. Mean acquisition performance time as a function of level of feedback, and trial block

The effect of levels of embedding on acquisition errors is shown in figure 8. In spite of different levels of embedding for a simple task, there is no clear-cut effect on error scores ($F = .22$; $df = 2,36$; $p > .05$). Significant variation in performance time is seen, however, in figure 9 ($F = 4.13$; $df = 2,36$; $p < .05$). Increasing levels of embeddedness clearly result in increasing performance time. What makes this result particularly interesting is that the number of task responses is constant across levels of embedding. Clear-cut interactions of embedding with other task parameters were not obtained.

Based on the preceding analyses, it was decided that a linear regression model would be appropriate for treatment of both acquisition error and time scores, since there were no striking interactions among task parameters which had to be taken into account. Consequently, in conducting these regression analyses there was no need to weight tasks differentially.

In an attempt to minimize potential confounding of results due simply to task length, however, acquisition error and time scores were transformed prior to analysis. The data selected for treatment were from the first (T_{1-2}), middle (T_{7-8}), and last (T_{13-15}) blocks of trials, these points being chosen to represent performance at early, intermediate, and later stages of acquisition. For each set of data, single variable regression analyses were conducted using number of task responses (TA) as the predictor variable. This procedure resulted in sets of residual criterion scores which were corrected for the effects of task length. While task length impacted upon performance, as noted in the preceding analyses, its effect was not of interest in the present study.

Six separate regression analyses were performed, one for each of the three time and three error criterion data sets. A step-wise regression procedure (Dixon, 1968) was employed with a maximum of three predictor variables being fitted. Standard values were employed for the F-level criteria for predictor variable inclusion or deletion. The results of the six analyses are summarized in table 1. Results are reported for three predictors. This conservative approach seemed warranted, given the rather small number of cases (n=20) involved. For each analysis, denoted by criterion data set, the multiple correlation coefficient (R) is reported together with the percentage of variance in the criterion accounted for (R^2). Also provided are the degrees of freedom (df) used in testing the significance of R and the resultant F-value. Finally, the specific indices included in each regression solution are listed. They appear from left to right in the order in which they were entered by the step-wise procedure.

As shown in table 1, even when the effect upon performance time due to number of responses (TA) is removed, significant multiple correlations between task indices and time are still obtained at all three acquisition stages. The important contributions of E% and C% to differences in performance time apparently reflect the extent to which superfluous equipment elements are encountered. As reported in a previous study (Wheaton and Mirabella, 1972) the extraneous equipment elements represented by such indices as E%, C% and D% apparently create a figure-ground problem which serves to retard performance time. The

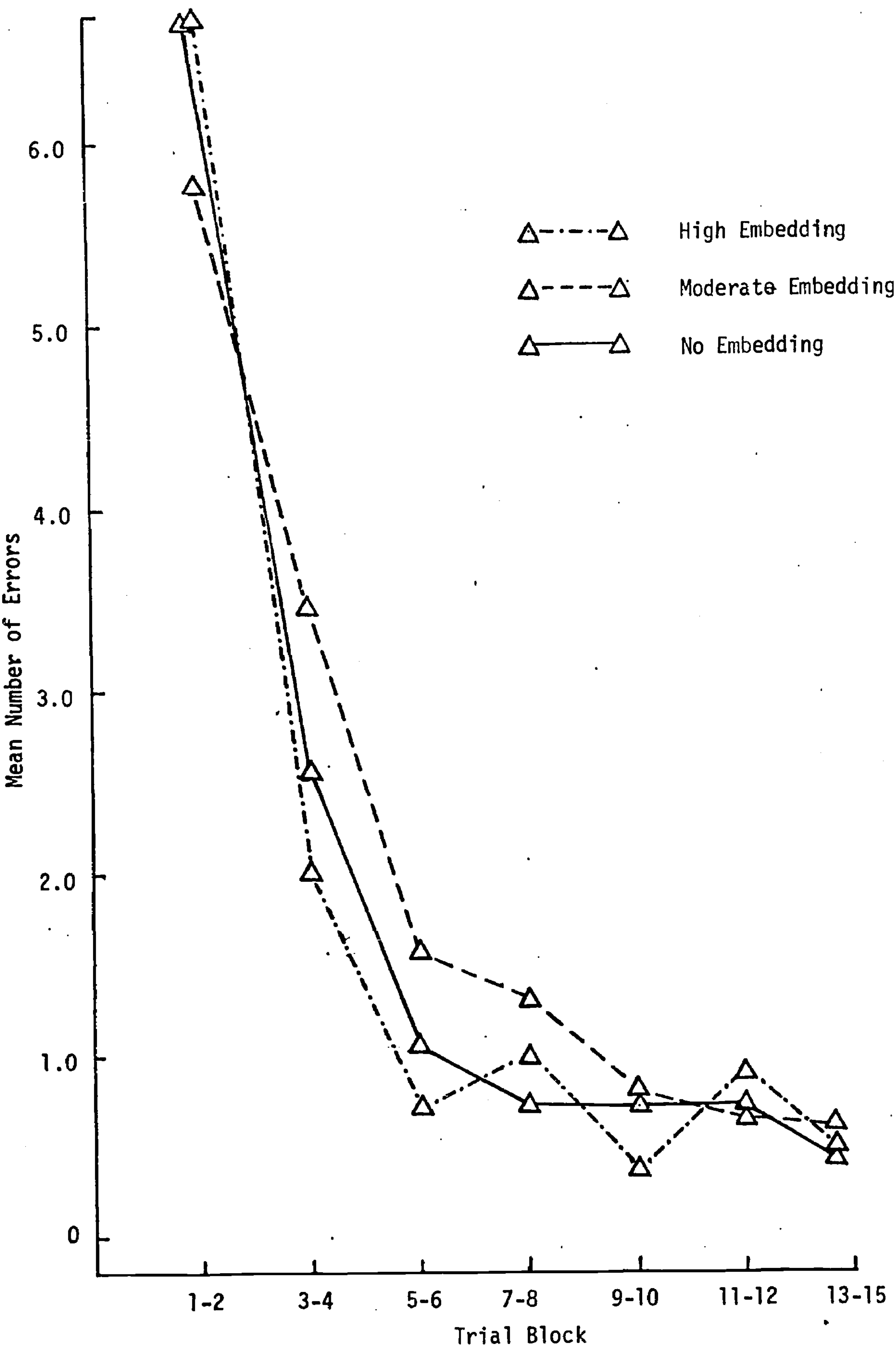


Figure 8. Mean acquisition errors as a function of level of embedding and trial block

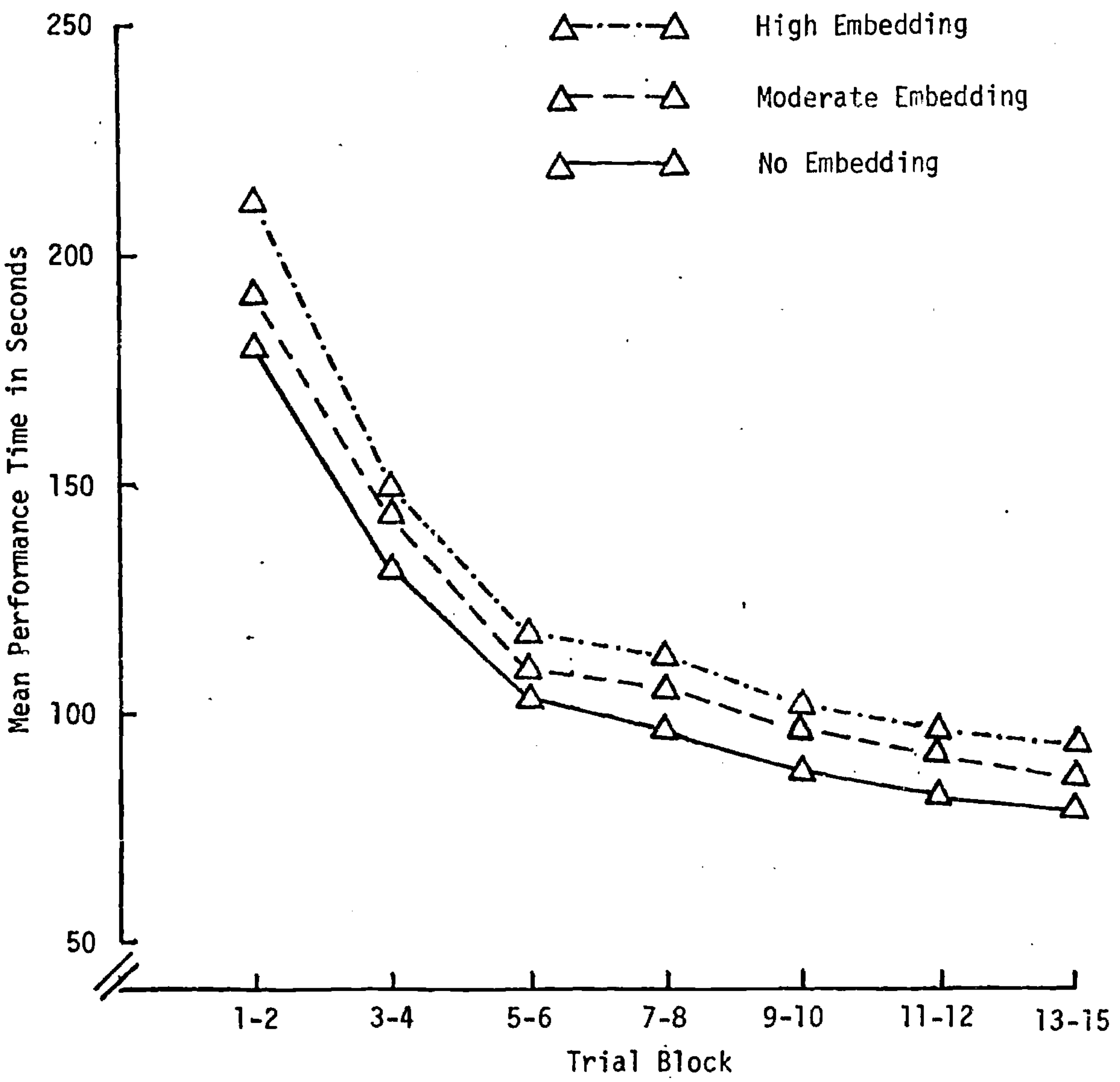


Figure 9. Mean acquisition performance time as a function of level of embedding, and trial block

TABLE 2: SUMMARY OF MULTIPLE REGRESSION ANALYSES OF RESIDUAL PERFORMANCE TIME AND ERRORS FOR FIRST, MIDDLE, AND LAST BLOCK OF ACQUISITION TRIALS

Criterion	R	R^2	df	F	Indices in order of selection by step-wise regression program
			Time Scores		
T_{1-2}	.693	.480	3, 16	4.92*	E%, DEI, CONT
T_{7-8}	.673	.453	3, 16	4.41*	C%, F%, INFO
T_{13-15}	.619	.383	3, 16	3.31†	C%, DEI, DISP
			Error Scores		
T_{1-2}	.474	.225	3, 16	1.55	E%, F%, D%
T_{7-8}	.670	.448	3, 16	4.33*	DEI, FBR, C%
T_{13-15}	.527	.278	3, 16	2.05	DEI, DISP, AA%

† $p. < .05$.

* $p. < .025$.

contribution of the DEI index to performance time is also of obvious importance, this rather complex index representing the ease with which an operator interacts with a particular set of displays and controls.

Findings with respect to error criterion scores are less dramatic. The only significant relationship occurs during the middle of acquisition. Here again, however, error rate is related to the goodness of information flow (DEI) associated with a given task. Generally, both sets of results continue to indicate that task indices of the type employed in the present study can be related to skill acquisition criteria.

The conservative nature of the analyses based on data corrected for TA can be appreciated by contrasting them with the raw score analyses shown in table 2. As shown in table 2, the multiple correlations for both time and error data are much higher when these data are analyzed in their raw form. More importantly, however, there is considerable overlap between both sets of analyses in terms of the task indices which relate most strongly to acquisition criteria. This overlap provides further support for the stability of the relationship between selected task characteristics and acquisition criteria.

TRANSFER. With respect to transfer data, only one of the task conditions employed in Phase II research was replicated during Phase III. Time and error transfer data obtained from these two research phases are presented in figures 10 and 11, respectively. In neither case is the main replication effect significant. In the case of performance time, however, there is a small but significant interaction between replications and trial blocks ($F = 3.99$; $df = 4,32$; $p < .025$). The small initial disparity in performance time disappears across blocks of trials. No such interaction was found between errors and trial blocks.

Transfer time and error measures were available for a sample of 15 different tasks, data for four of which were carried over from Phase II research. Prior to regression analysis, these data, like the acquisition data reported upon earlier, were examined in a series of linear contrasts. The purpose of these preliminary analyses was to determine the appropriateness of an additive linear model when attempting to relate task indices to transfer criteria.

The main effects of complexity, feedback, and trial block were found to impact upon transfer performance as expected. The interactions among these variables are presented in figures 12 through 14. In interpreting these findings it should be recalled that the data reflect scores on the second or transfer task (Ma). As shown in figure 12, the impact of task complexity of the acquisition task, on transfer task errors, interacts with the presence or absence of feedback in the first task and trial block on the transfer task ($F = 2.15$; $df = 8,96$; $p < .05$). Transfer from the more complex device (Ca) is better than transfer from the less complex device (Sa), given that the "critical" feature of feedback is present. Presence or absence of feedback during training has its most marked effect on transfer for complex tasks, its smallest effect for simple tasks, and an intermediate effect for the medium task. These differences tend to diminish over trial blocks although they are still prevalent on the last transfer trial (T_{9-10}). The transfer time

TABLE 3. SUMMARY OF MULTIPLE REGRESSION ANALYSES OF UNADJUSTED TIME AND ERROR SCORES FOR FIRST, MIDDLE, AND LAST BLOCK OF ACQUISITION TRIALS

Criterion	R	R^2	df	F	Indices in order of selection by step-wise regression program
			Time Scores		
T_{1-2}	.874	.764	3, 16	17.30**	DEI, FBR, E%
T_{7-8}	.908	.825	3, 16	25.15**	DEI, E, C%
T_{13-15}	.920	.847	3, 16	29.60**	TA, DEI, C%
			Error Scores		
T_{1-2}	.669	.448	3, 16	4.32†	DEI, LV, E%
T_{7-8}	.809	.655	3, 16	10.13**	DEI, CRPS, FBR
T_{13-15}	.766	.586	3, 16	7.56**	CRPS, AA%, DEI

† $p < .05$.
** $p < .01$.

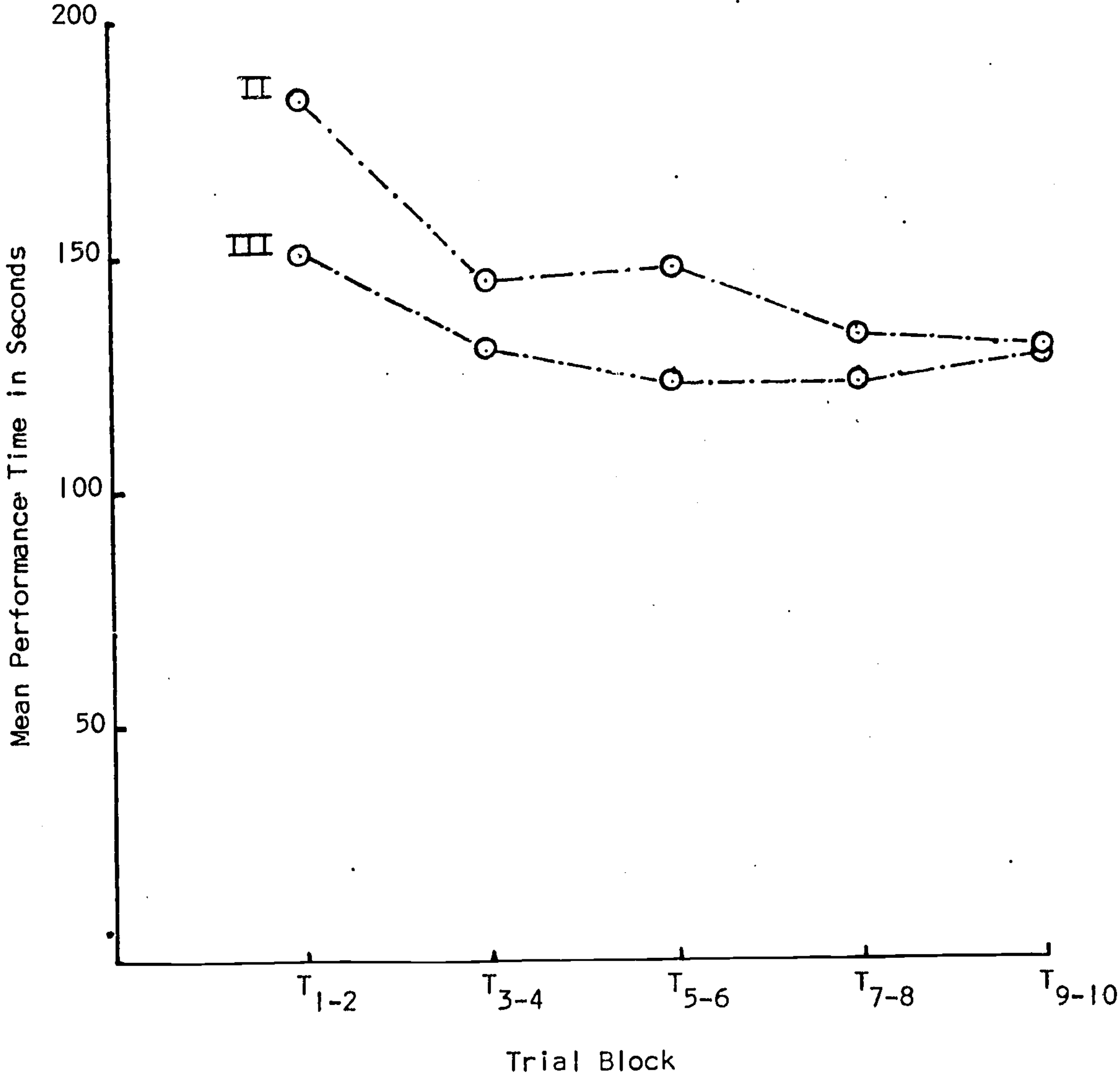


Figure 10. Mean performance time as a function of trial block during transfer of training to task Ma, following acquisition on Task SEca (Phase II and Phase III data compared)

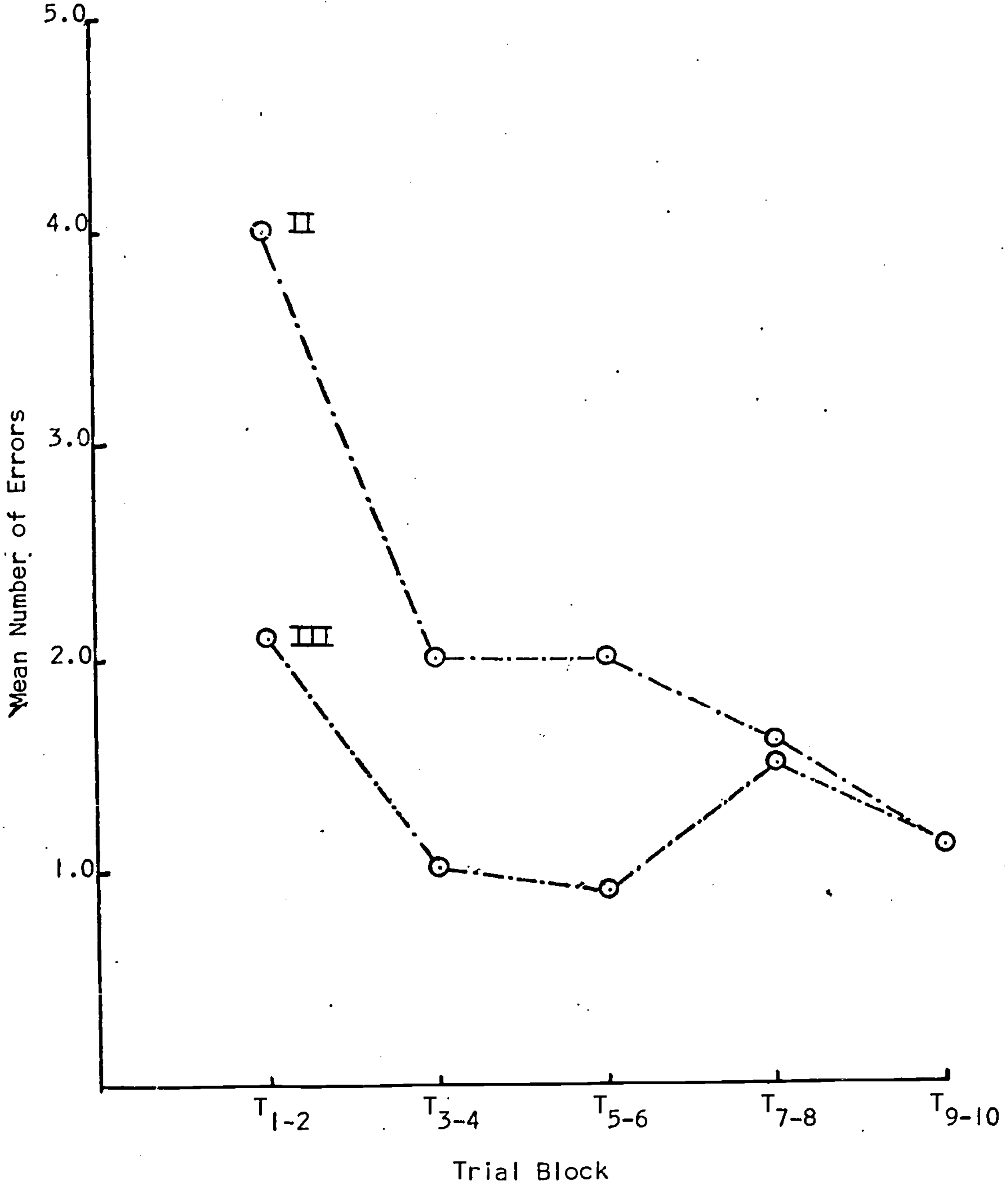


Figure 11. Mean number of errors as a function of trial block during transfer of training to task Ma following acquisition on SEca (Phase II and Phase III data compared)

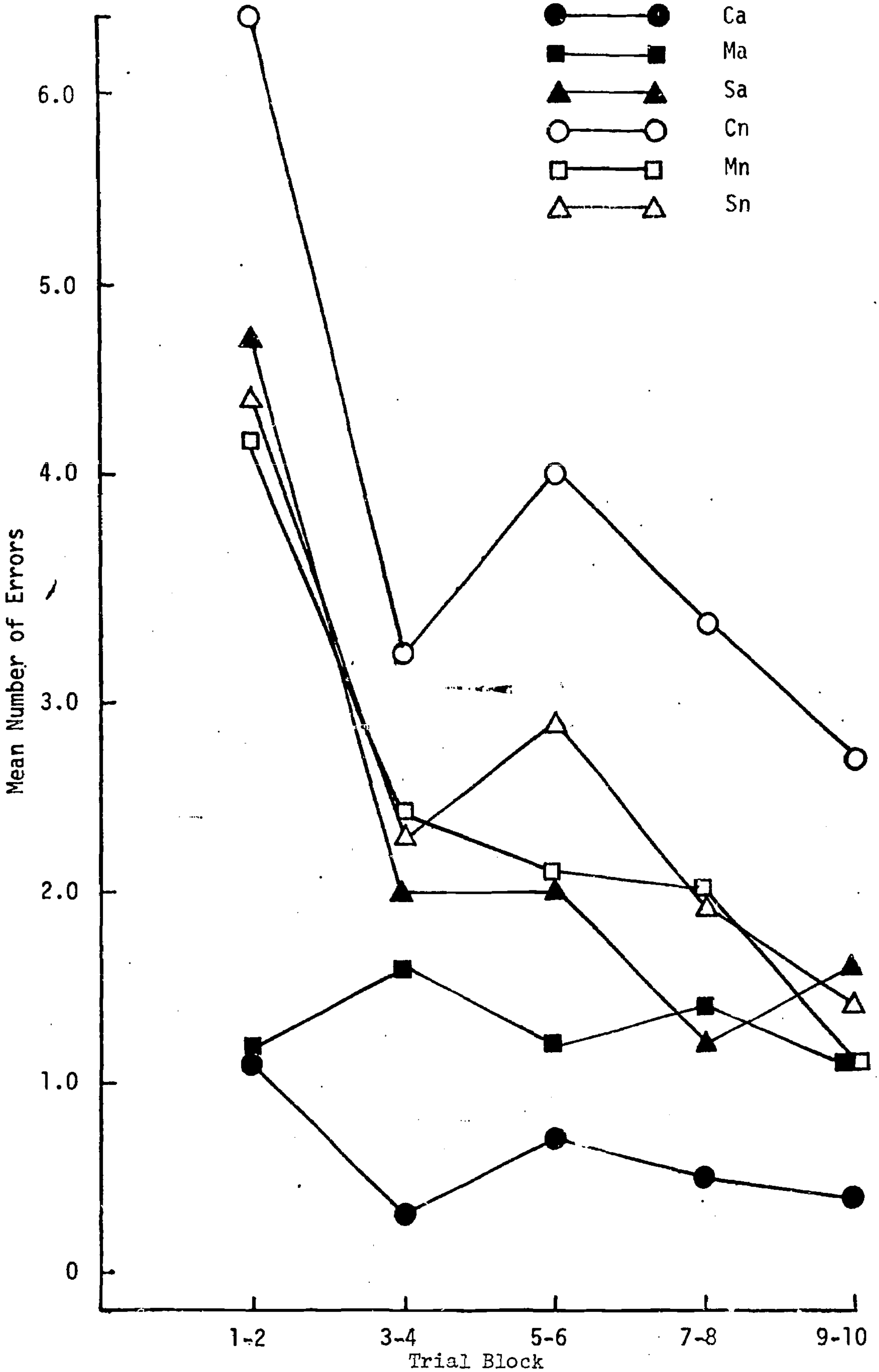


Figure 12. Mean errors during transfer as a function of acquisition task complexity, amount of feedback, and trial block

data shown in figure 13 are subject to a similar complex interaction of task complexity, feedback, and trial ($F = 3.18$; $df = 8,96$; $p < .005$).

Embedding, while not significant as a main effect, did interact with feedback and trials for both error ($F = 2.30$; $df = 16,144$; $p < .01$) and time ($F = 1.97$; $df = 16,144$; $p < .025$) scores during transfer. Particularly interesting is the general positive effect which embedding of the training task has on the accuracy of transfer performance (figure 14). Increasing embeddedness shows evidence of increasingly better transfer, i.e., performing a simple task embedded in a more complex console facilitates transfer to a more complex task.

Considered collectively, the results of these preliminary analyses indicated the presence of a number of complex interactions among task parameters on transfer criteria. These findings suggested that while an additive linear regression model could be used in investigating acquisition data, it would not be particularly powerful in dealing with transfer data. Accordingly, an attempt was made to differentially weight task parameters, thereby reducing nonlinearities in the transfer data. The weights were derived from the facts that: (1) disruptive effects of no feedback diminish as task complexity decreases; and (2) partial feedback for simple tasks is more disruptive than the no-feedback condition.

Based upon these generalizations and as a tentative approximation, a set of ordinal weights was applied to the DEI index. This index was chosen for weighting because it seemed to be the single index most representative of task complexity, the dimension underlying many of the interactions. The weights were applied only to non-embedded tasks as follows: Cn, 3; Mn, 2; Ss, 1.5; Sn, 1. The DEI's of all other tasks received a weight of 1. These weights followed from consideration of points (1) and (2) above.

Six regression analyses were performed on the raw transfer data. Since a single transfer task had been used, there was no need to correct error or time data for task length. The dependent measures consisted of error and time data obtained at an early point (T_{1-2}), an intermediate point (T_{5-6}), and later on (T_{9-10}) during transfer. The independent or predictor measures consisted of the absolute difference scores ($\triangle$) between the acquisition task and the transfer task for each of 14 task indices. (See Appendix A.) As previously noted, a weighted DEI index was used in these analyses.

As shown in table 3, significant multiple correlations are obtained between task indices and both time and error measures at each stage of transfer. Within the analyses concerned with performance time, there is an obvious consistency in the set of predictors relating to the criterion at each stage of transfer. The differences (between acquisition and transfer tasks) in the number of displays ($\triangle$DISP), the percentage of controls used ($\triangle$C%), and the weighted Display Evaluative Index ($\triangle$DEIW) bear strong relationships to the criterion at each point. The predictors of errors during transfer are not as consistent over trial blocks, with the exception, perhaps, of the weighted DEI measure and the equipment element index ($\triangle$E).

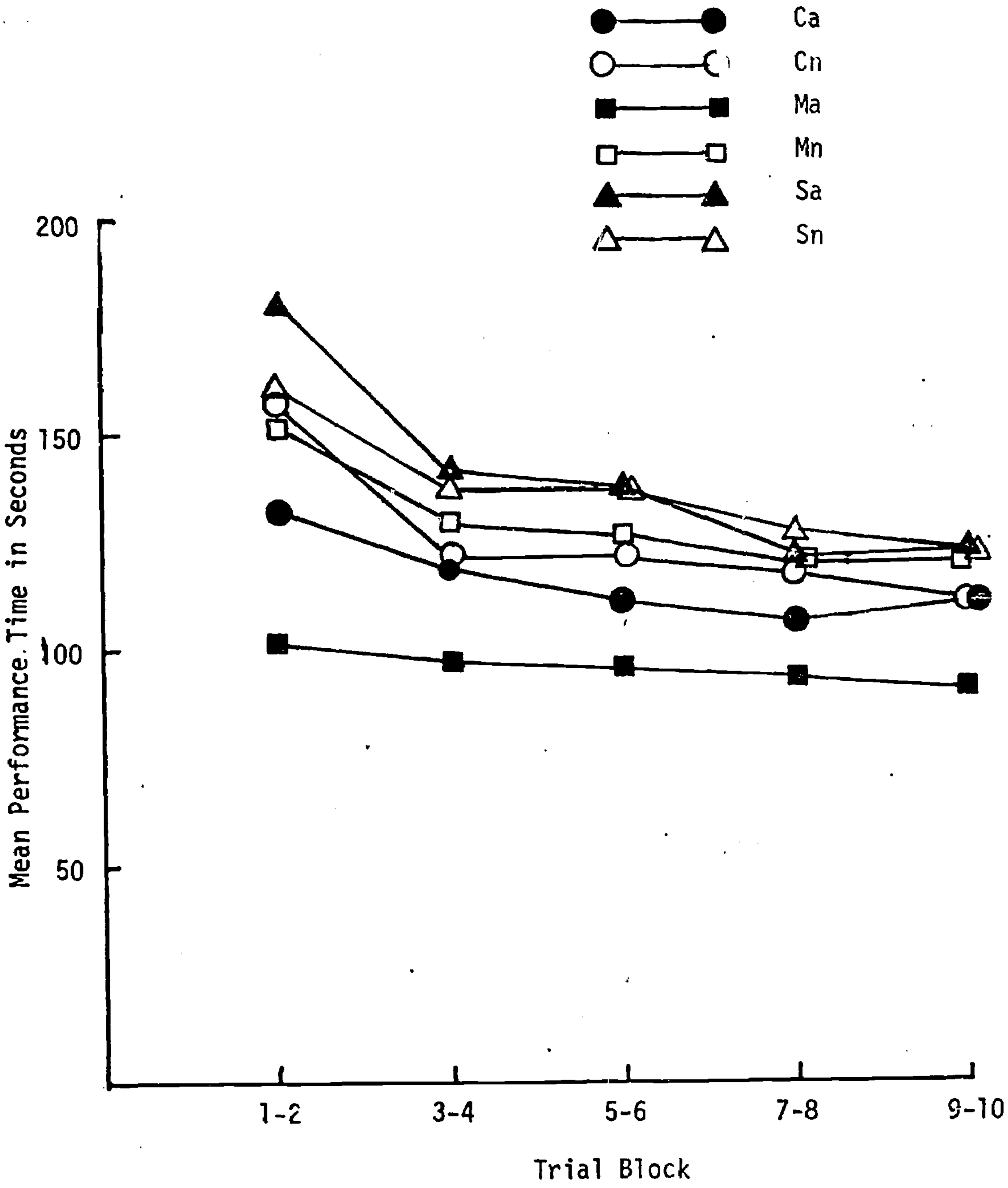


Figure 13. Mean time during transfer as a function of acquisition task complexity, amount of feedback, and trial block

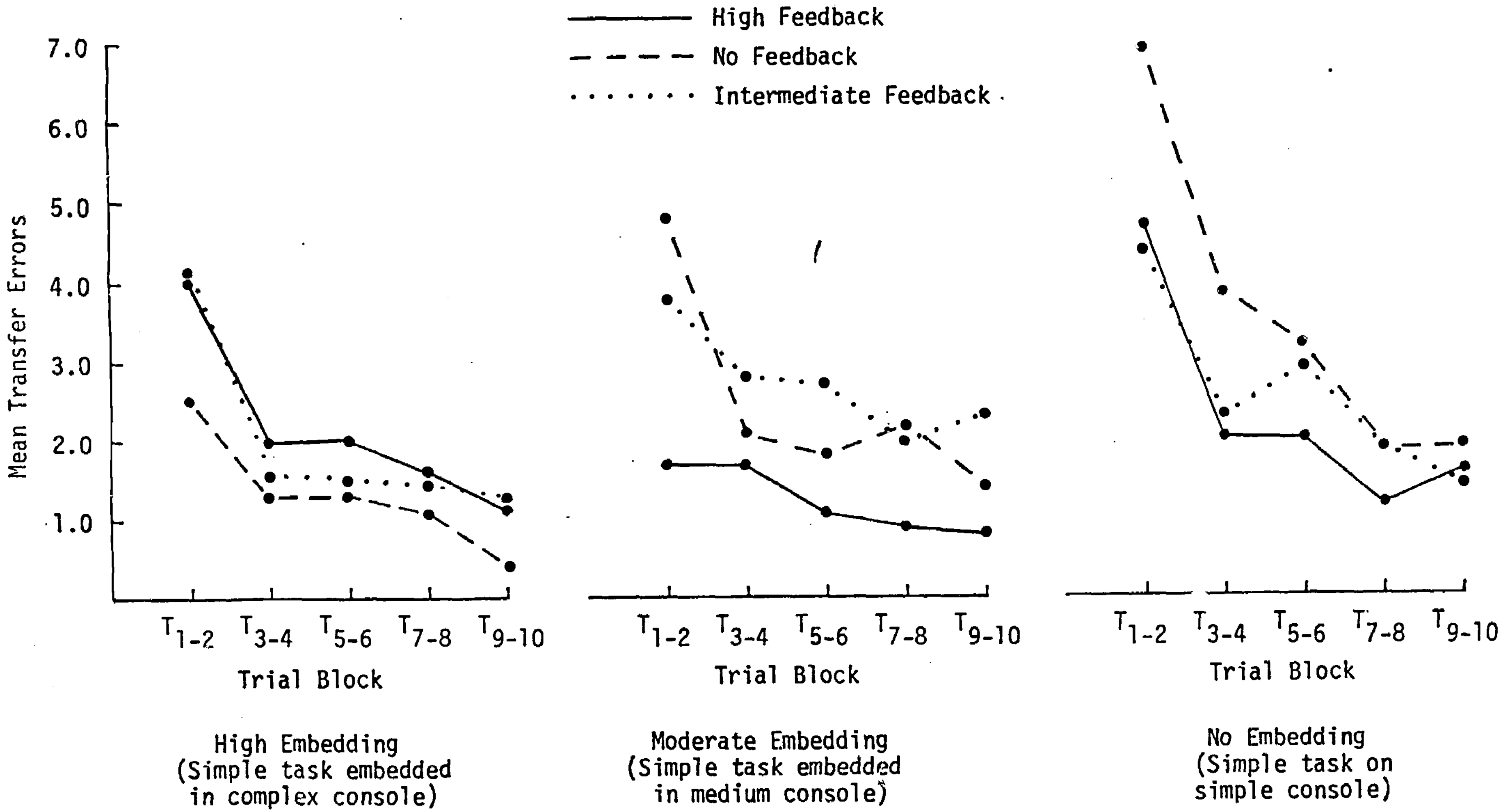


Figure 14. Mean errors during transfer as a function of embedding, feedback, and trial block

TABLE 4: MULTIPLE REGRESSION ANALYSES USING DIFFERENCE SCORES TO PREDICT RAW TIME AND ERROR SCORES FOR FIRST, MIDDLE, AND LAST BLOCK OF TRANSFER TRIALS

(WEIGHTED DEI INDEX)

Criterion	R	R^2	df	F	Indices* in order of selection by step-wise regression program
			Time Scores		
T_{1-2}	.751	.564	3, 11	4.75†	ΔDISP, ΔC%, ΔDEIW
T_{5-6}	.771	.595	3, 11	5.39†	ΔDISP, ΔC%, ΔDEIW
T_{9-10}	.805	.648	3, 11	6.76*	ΔDISP, ΔC%, ΔD%
			Error Scores		
T_{1-2}	.890	.793	3, 11	14.03**	ΔDEIW, ΔINFO, ΔFBR
T_{5-6}	.914	.836	3, 11	18.67**	ΔDEIW, ΔE, ΔF%
T_{9-10}	.824	.679	3, 11	7.75*	ΔDEIW, ΔE, ΔD%

* Indices represent absolute differences between acquisition and transfer tasks.

†p. < .025.

*p. < .01.

**p. < .001.

For the sake of comparison, additional regression analyses based upon alternative sets of predictors are presented in tables 4 and 5. The regression analyses shown in table 4 are based on the same set of predictors as used in table 3, with the exception of the DEI index, which appears in its unweighted form. The two sets of analyses are quite similar with respect to the pattern of predictors entered into each solution. Generally, however, slightly larger multiple correlation coefficients are obtained when the weighted (table 3) as opposed to the unweighted (table 4) DEI index is used.

As shown in table 5, strong multiple correlation coefficients are also obtained when the actual index values of the various acquisition tasks are used as the predictor values. The resultant patterns of predictors are somewhat less consistent over trial blocks within the time or error analyses relative to those patterns shown in tables 3 and 4. Also of interest is the difference in the magnitude of the multiple correlation coefficients obtained when the predictors are based on actual task index values (table 5) or difference values (tables 3 and 4). The use of actual task index values leads to higher coefficients for time measures early during transfer. Later for time scores, however, and generally throughout the transfer session for error scores, the use of absolute difference (| transfer task minus acquisition task |) values for the various indices results in higher regression coefficients.

To summarize, it has been possible to demonstrate with this series of experiments that variations in quantitative task indices can be related significantly and consistently to trainee performance. It should be emphasized, however, that while the focus of the research just described was upon trainee task variables, it is recognized that this class of variables is not the only one which impacts upon device effectiveness. Training method, including device utilization, may be as potent, if not more so. To investigate these issues, principally the interaction between task complexity as measured by the task indices, and method of training, a second experiment was conducted. The results are presented below.

STUDY 2: INTERACTION BETWEEN TASK CHARACTERISTICS AND TRAINING METHODS

Analyses were conducted to examine the effects upon acquisition and transfer criteria of variations in task characteristics and training methods. The data were analyzed using three designs which permitted examination of the interactions among these classes of variables (Appendix C).

In preparing for these analyses zero-order correlations were computed between subjects' acquisition and transfer time and error scores on the one hand, and associative memory test scores on the other hand. The latter measures were obtained with the expectation that they might serve as useful covariates, by means of which differences in performance which were not functions of the experimental treatments per se might be controlled for. The correlations between the covariate and variate measures, however, were essentially zero, indicating that a covariate adjustment of the performance data would have little utility. Accordingly, analyses of variance were conducted, the major results of

TABLE 5 : MULTIPLE REGRESSION ANALYSES USING DIFFERENCE SCORES TO PREDICT RAW TIME AND ERROR SCORES FOR FIRST, MIDDLE, AND LAST BLOCK OF TRANSFER TRIALS

(UNWEIGHTED DEI INDEX)

Criterion	R	R^2	df	F	Indices* in order of selection by step-wise regression program
			Time Scores		
T_{1-2}	.717	.514	3, 11	3.87†	ΔDISP, ΔC%, ΔFBR
T_{5-6}	.747	.559	3, 11	4.64†	ΔDISP, ΔC%, ΔDEI
T_{9-10}	.805	.648	3, 11	6.76*	ΔDISP, ΔC%, ΔD%
			Error Scores		
T_{1-2}	.734	.539	3, 11	4.29†	ΔDEI, ΔE, ΔDISP
T_{5-6}	.810	.656	3, 11	6.99*	ΔDEI, ΔE, ΔDISP
T_{9-10}	.794	.630	3, 11	6.24*	ΔDEI, ΔE, ΔDISP

* Indices represent absolute differences between acquisition and transfer tasks.

† $p. < .05$.

*$p. < .01$.

TABLE 6: MULTIPLE REGRESSION ANALYSES USING ACQUISITION TASK INDEX VALUES TO PREDICT RAW TIME AND ERROR SCORES FOR FIRST, MIDDLE, AND LAST BLOCK OF TRANSFER TRIALS

(UNWEIGHTED DEI INDEX)

Criterion	R.	R^2	df	F	Indices* in order of selection by step-wise regression program
			Time Scores		
T_{1-2}	.835	.698	3, 11	8.46*	E, INFO, F%
T_{5-6}	.820	.672	3, 11	7.53*	E, INFO, F%
T_{9-10}	.728	.530	3, 11	4.14†	E, TA, C%
			Error Scores		
T_{1-2}	.749	.560	3, 11	4.67†	FBR, D%, AA%
T_{5-6}	.779	.607	3, 11	5.65†	FBR, D%, INFO
T_{9-10}	.661	.437	3, 11	2.84	FBR, INFO, E%

* Indices represent values on acquisition tasks.

†$p. < .05$.

*$p. < .01$.

which are presented in figures 15-19 for both acquisition and transfer data.

ACQUISITION. The impact of task complexity on acquisition criteria was similar to that reported earlier for the transfer of training study. Significant interactions between task complexity and trial blocks were obtained for acquisition errors ($F = 4.95$; df = 6,144; $p < .01$) and acquisition time ($F = 6.57$; df = 6,144; $p < .01$). The interactions arose from a convergence in "simple" and "complex" task performance over trial blocks. For example, on the first trial block a mean of 12.0 errors occurred on the "complex" task relative to 7.2 errors on the "simple" task. On the last acquisition trial more errors were still associated with the "complex" task (1.4), but the difference between the two was smaller (i.e., mean errors on the simple task = 0.3). Similar patterns were obtained for time measures.

Task embedding had no significant effect upon acquisition performance for either error ($F = .52$; df = 2,36; $p > .05$) or time ($F = .58$; df = 2,36; $p > .05$) scores. The lack of an error effect is comparable to Study 1 findings. On the other hand, the time effect found in Study 1 was not obtained, a result which is attributable, perhaps, to the different tasks used in the two studies.

Finally, there is evidence that training method affects the number of errors made during acquisition ($F = 3.53$; df = 2,24; $p < .05$). Most errors occur when the cold-panel method is used (mean = 3.71 errors). The hot-panel and pictorial methods are comparable, producing fewer errors (pictorial mean = 2.43 errors; hot-panel mean = 2.39 errors).

A more complete presentation of these results, however, is given in figure 15, where errors are shown as a function of the interaction between task complexity and training method. This interaction approached significance ($F = 3.02$; df = 2,24; $p \approx .07$), and tended to indicate that the relative inferiority of the cold-panel approach holds only for the complex task situation. Training method did not influence performance time during acquisition.

TRANSFER. Training task complexity has a significant impact on error scores during transfer ($F = 4.75$; df = 1,24; $p < .05$). Fewer errors (mean = 1.09) occur following acquisition training on a task more complex than the transfer task, and relatively more (mean = 1.89) after acquisition training on a task simpler than the transfer situation. These results are similar to those reported earlier for Study 1, when both of these tasks possessed a high level of feedback.

Time scores during transfer are a function of an interaction between acquisition task complexity and trial block ($F = 4.25$; df = 4,96; $p < .01$). The initial spread between simple and complex tasks and their subsequent convergence over trials are shown in figure 16. Of particular interest is the general facilitation in transfer performance time on a task of medium complexity, having practiced on a more complex task. These results are highly similar to those reported earlier in figure 13 for tasks possessing feedback.

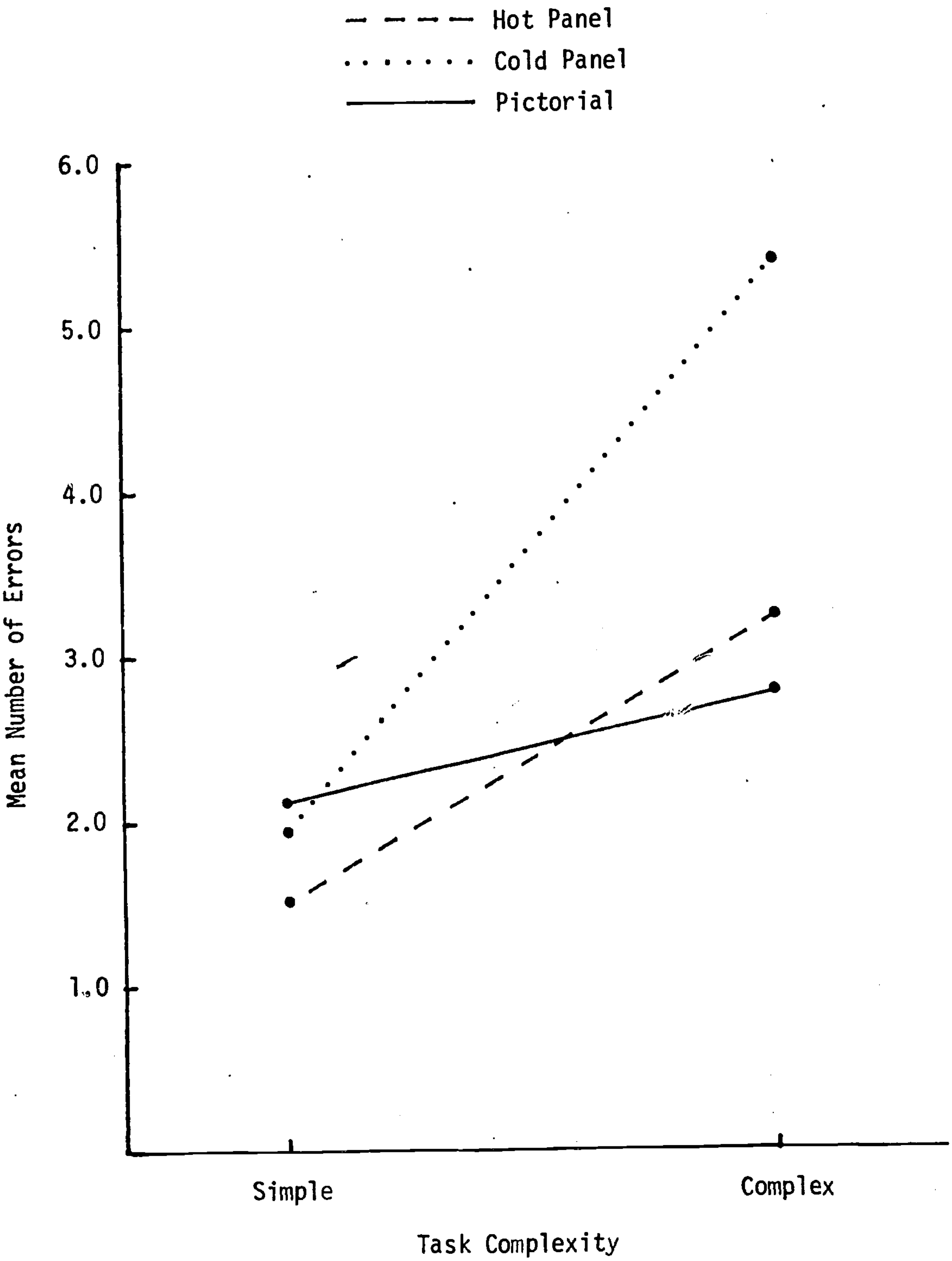


Figure 15. Mean acquisition errors as a function of task complexity and training method

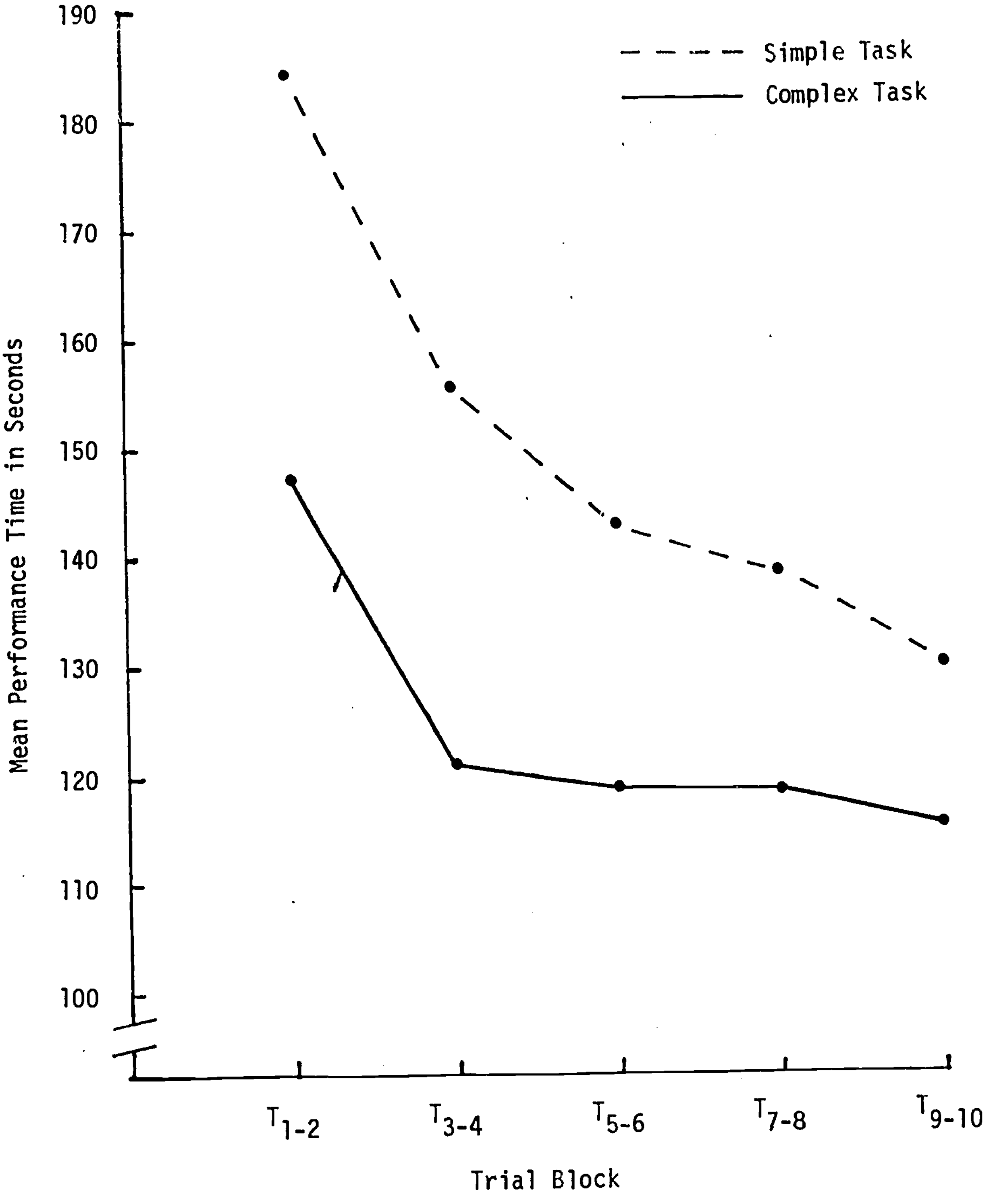


Figure 16. Mean transfer time as a function of training task complexity and trial block

Unlike the findings presented for Study 1, Study 2 data suggested that neither embedding per se nor the level of embedding has any main or interactive effect on the errors made during transfer. In Study 1, embedding interacted with level of feedback and trial block to affect error rate. With respect to time scores, however, embedding of the acquisition task interacts in a complex manner with training method and trials to determine performance time during transfer ($F = 2.58$; $df = 8,192$; $p < .01$). This relationship is shown in figure 17. Relatively faster performance time occurs after training on the hot panel, but the advantage of this method over the other two is moderated by embedding of the acquisition task.

The results just presented are the only case in which method of training interacts with a task parameter to affect transfer error or time. Consistently, however, training method interacts with trials to determine performance during transfer. A significant training method by trials interaction ($F = 2.11$; $df = 8,192$; $p < .05$) is shown in figure 18 for transfer errors. The relative superiority of training on the hot panel early in transfer decreases over time. By the end of the transfer period, the three methods are virtually the same in terms of error rates. A significant training method by trial interaction for transfer performance time ($F = 2.60$; $df = 8,144$; $p < .01$) is shown in figure 19 for the simple task. Notice that the difference in performance time between the hot-panel and cold-panel groups is maintained across the entire transfer period, while the pictorial group, after an initial retardation relative to the hot-panel group, rapidly converges with it.

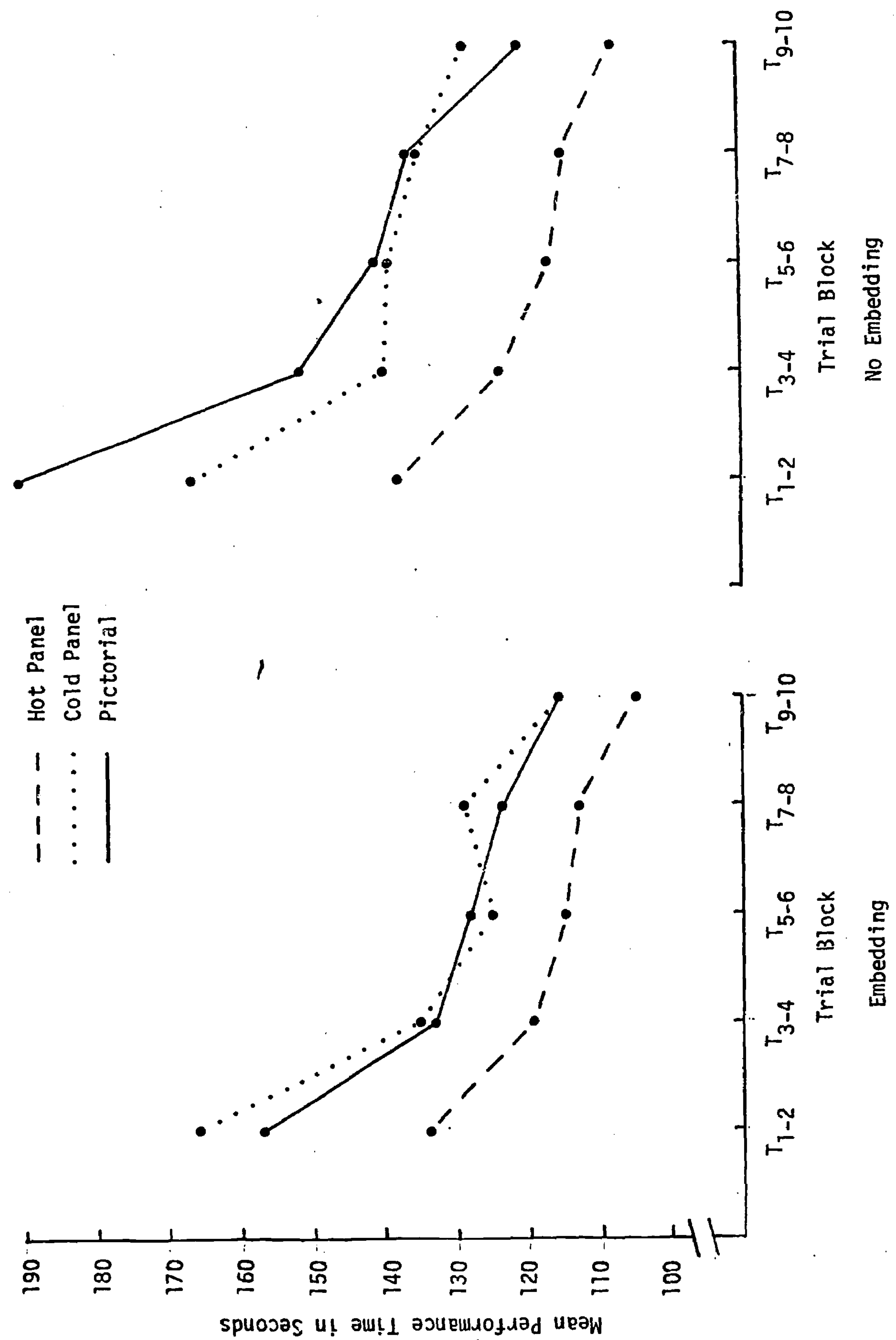


Figure 17. Performance time during transfer as a function of embedding, training method, and trial block

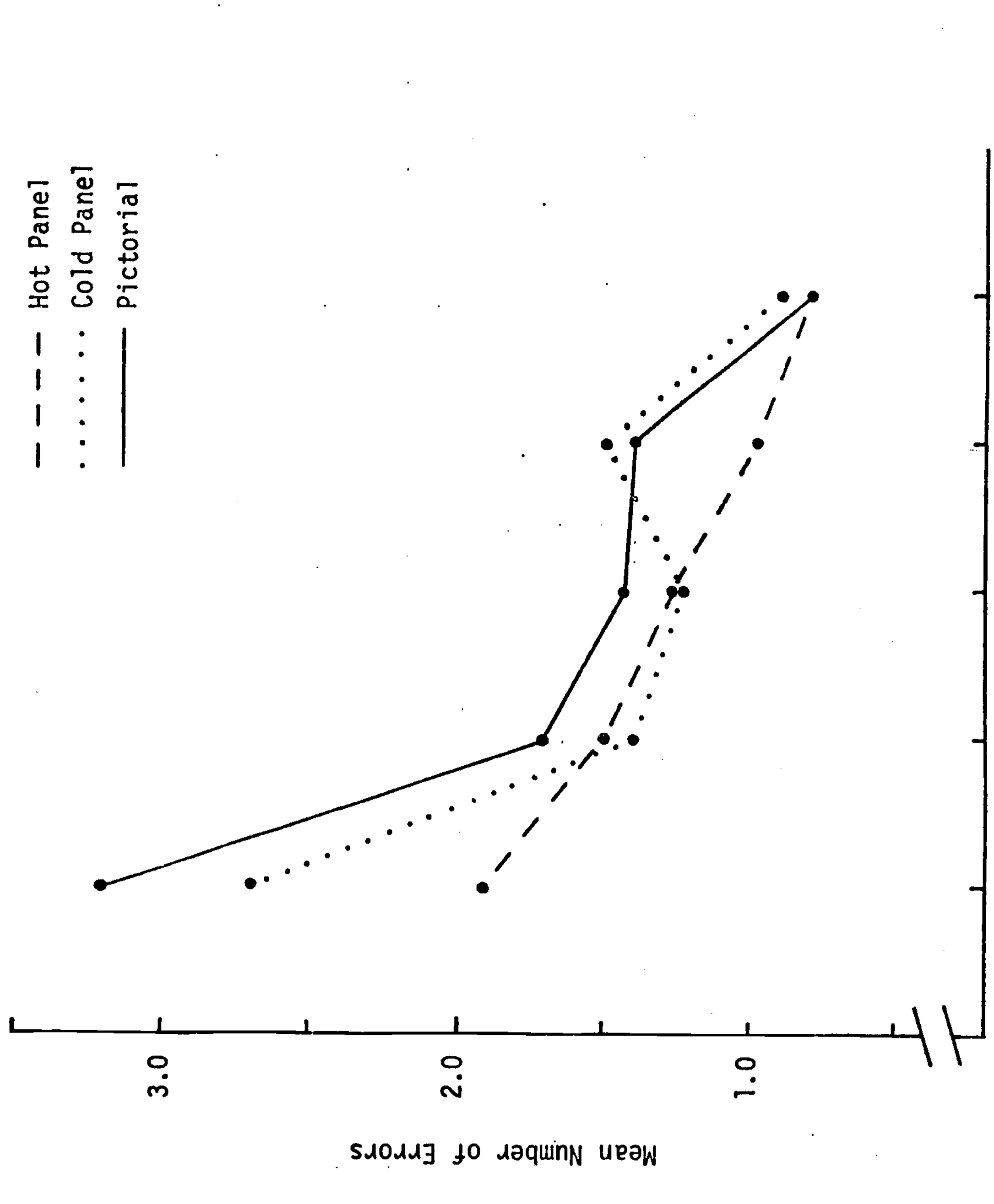


Figure 18. Mean transfer error as a function of training method and trial block

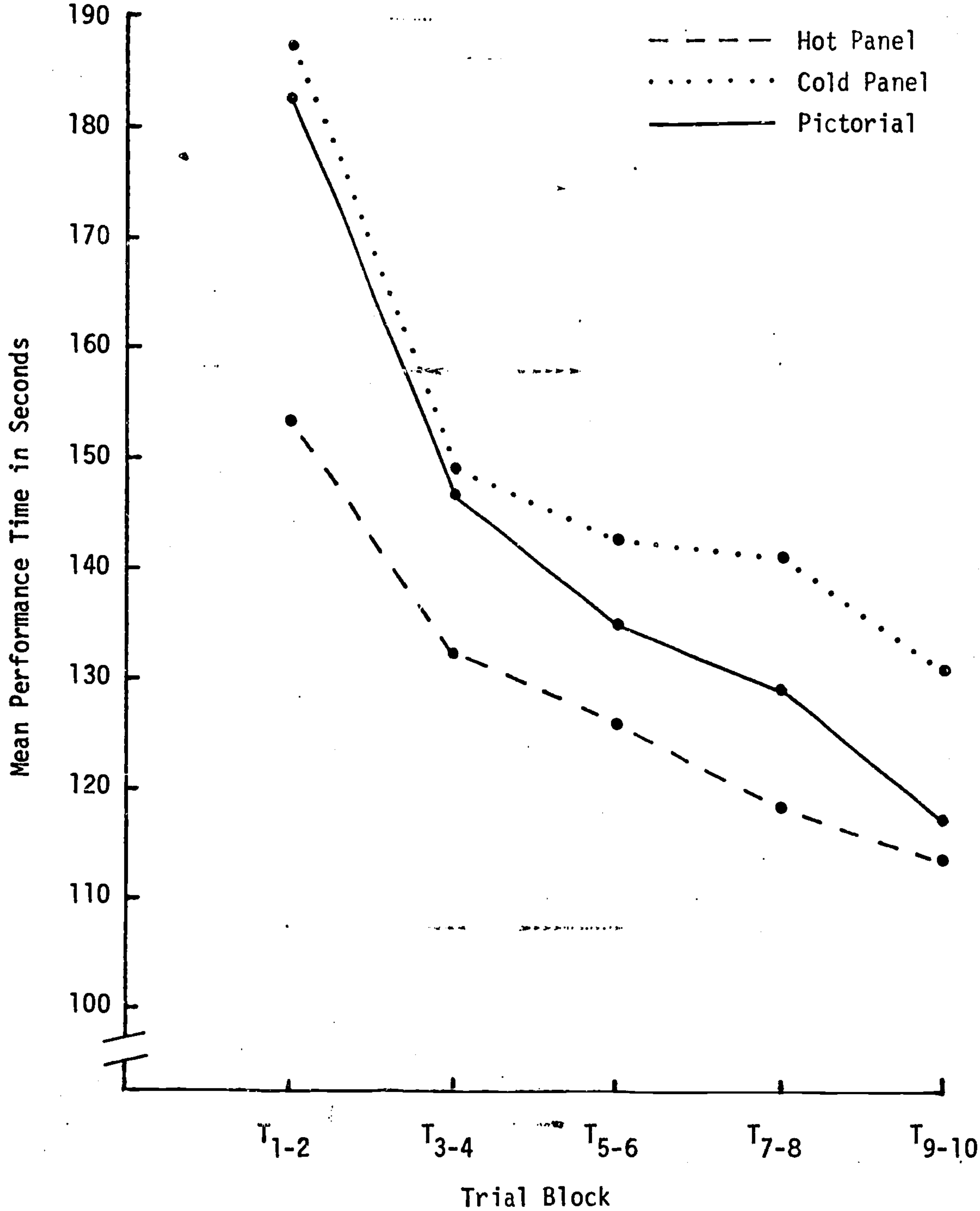


Figure 19. Mean transfer time as a function of training method and trial block

SECTION IV

DISCUSSION

In this section, the results detailed in Section III are reviewed for Studies 1 and 2 separately. Their implications for task quantification and performance prediction are then discussed. Finally, major conclusions and implications for further development and use of the predictive methodology are drawn.

PREDICTION OF ACQUISITION

In many respects the results of Study 1 corroborated those obtained in Phase II (Wheaton and Mirabella, 1972). Consistently large and intuitively systematic variations in performance were obtained as a function of task/trainer configuration. Once again these variations persisted even when the effects of task length were removed.

Further indication of the reliability of the earlier results was obtained when a number of Phase II tasks were replicated and found to yield comparable performance curves. The strength of this stability can be better appreciated if it is recalled that sample size per task examined was very small, a situation in which the likelihood of distortions caused by a few aberrant scores is high.

The predictive power of the indices for skill acquisition was upheld, with multiple correlations substantially the same as found in Phase II. The pattern of predictors changed somewhat in Phase III, but this is not unreasonable since the number of cases entering the regression analysis nearly doubled and, moreover, the number of predictors utilized was reduced from seventeen to fourteen. A more stable analysis would be expected in this case, and this could very well be accompanied by a somewhat different selection of optimum predictors. Accordingly, the Phase III predictors for acquisition are to be preferred to those obtained in Phase II. For example, DEI enters prominently in Phase III among the predictors of both time and error scores. It did not appear at all in Phase II analyses. Its appearance in Phase III, however, is consistent with the greater variety of acquisition tasks since descriptively it is the most inclusive of all 14 indices.

A number of indices were common to the acquisition analyses of Phases II and III. In both phases, for example, E% was predictive of both errors and time early in acquisition. Thus, the importance of task embeddedness, as reflected by the E% index, was corroborated. Note here that the relationship between E% and performance is inverse. That is, both errors and task completion times are reduced as E% increases. In other words, as task embedding decreases, performance during acquisition improves.

As in Phase II, the pattern of predictors was shown to vary across criterion measures and across time blocks within criterion measures. Thus, a simple figure-of-merit approach to device evaluation was not supported, at least in terms of acquisition performance.

PREDICTION OF TRANSFER

The suggestion in Phase II that the index battery might be extendable to transfer of training criteria was upheld by the transfer analysis of Phase III. Using task characteristic difference scores, very substantial multiple-correlation coefficients were obtained for both performance time and error, and across time blocks within criteria. These coefficients were considerably stronger than for acquisition. Furthermore, consistency of predictor sets was markedly greater, not only within criteria, but across criteria as well. DEI again was prominently represented, an encouraging finding since DEI is the most inclusive index in the battery. DEI was particularly in evidence for error criteria, along with number of displays and controls (E) and number of displays (DISP). That is, decreasing differences between the acquisition and transfer tasks on the DEI, E, and DISP indices were related to decreasing time and error scores during transfer. The improved consistency found in these data, in contrast to the acquisition analyses, provides correspondingly greater encouragement for a figure-of-merit approach when transfer of training criteria are employed.

The validity of the difference scores as predictors of performance during transfer has particular significance. One of the criticisms levied against a task-similarity model of transfer of training is that similarity is typically unquantifiable except for very simple laboratory tasks (e.g., pitch discrimination). The current results provide an instance in which it was possible to quantify similarity for a surrogate "real-world" task and to predict performance with very high validity. High validity was obtained notwithstanding an interaction between task complexity and feedback, one of the underlying parameters used to manipulate DEI. In the preliminary linear contrasts which preceded regression analysis of the transfer data, it was found that absence of feedback lights had a disruptive influence upon performance. The disruption was greater for the complex than for the simple task. This interaction had the effect of transforming DEI into a nonlinear variable vis a vis performance error, thus reducing its power for linear regression. It was for this reason that a linearizing transformation of DEI was attempted. Substantial increases in the multiple correlations resulted from this transformation as was shown in the contrasting multiple-correlation tables. An alternative treatment would have been to develop two predictor equations, one including feedback cases, the other including no-feedback cases. However, sample size was too small to permit this approach.

The significance of the foregoing exercise goes beyond the feedback issue, since obviously no training device designer is going to opt for the removal of status indicators from a trainer console. But to the extent that analogous effects can be identified and appropriately weighted by the user of the indices, their predictive power will be increased. Even with some index interactions, however, the data suggest that a linear regression model will still provide good predictability of transfer of training criteria.

STUDY OF TRAINING METHODS AS A FUNCTION OF TASK COMPLEXITY

In addition to its utility as a predictive tool, another potential value of task quantification is that it can aid significantly in studying interactions among the different classes of parameters which may impact upon device effectiveness. If, for example, one were interested in understanding how task complexity and training methods intersected, it would be important to sample tasks over a broad range of complexity levels. A quantification methodology can help insure that such a range is covered and that the tasks studied do, in fact, differ significantly. Study 2 was designed primarily as a demonstration of how the indices could be applied to such a purpose.

The specific hypothesis of Study 2 was that the effectiveness of dynamic procedural training versus static training would depend upon task complexity as differentiated by the quantitative task indices. The characteristic conclusion of studies of procedural training has been that dynamic training is not cost-effective; namely. that acquisition of skills and transfer to operational contexts are essentially as good when mock-ups are used for training (Grimsley, 1969; Prophet and Boyd, 1970; Bernstein and Gonzalez, 1971).

The results of Study 2 provide some support for the hypothesis of an interaction between task parameters and method of training. During acquisition, training method appeared to have a differential effect for the complex task, with cold-panel presentation generating more errors than either pictorial or hot-panel presentation. Clearer support for an interaction is found in the transfer data where presence or absence of task embeddedness generated a differential performance effect for training methods. Dynamic presentation led to consistently faster performance across transfer blocks than either cold or pictorial presentation. Its superiority, however, was greater under the no-embedding condition.

Results of Study 2 were otherwise consistent with those of earlier studies. For example, the training method by trials interaction found for transfer was also reported by Bernstein and Gonzalez,(1971). In both studies an initial advantage of dynamic training, particularly in contrast to the pictorial method of training, rapidly dissipated.

The failure to generate more decisive data on the methods-by-task interaction may in part be due to the difficulty in controlling individual differences sufficiently. The covariate data (associative memory tests) which were collected in an effort to reduce error variance proved ineffective and could not be used for covariance analysis, as originally planned.

The potential significance of task quantification for studying interactions among major classes of variables is worth pursuing further. The alternative which has commonly been employed, for lack of a quantitative taxonomy, is to select tasks on an intuitive basis, and this is simply not satisfactory.

APPLICATION OF THE INDICES

Use of the indices (Appendix D) would be fairly straightforward if the particular beta weights emerging from Phase III were to be employed. These weights are presented in Appendix E. They can be applied directly to the raw task index values which would result from the analysis of two or more prototype devices. The resulting predicted performance values would then provide a basis for at least ordinal comparison of the prototypes.

At this level, the indices could be employed as one of several tools to support the training expert's evaluation of alternative prototype devices. They might be employed, for example, to corroborate or question judgments already established by other means.

More rigorous and confident use, however, requires cross validation on actual training devices. At least one reason for this requirement is that the range of index values employed in these researches was notably smaller than the range which would be found for field apparatus. For example, DEI ranged from approximately 5 to 20 in the laboratory effort. Values obtained on sonar trainers in the field ranged from approximately 3 to 65. While this increased range should maintain or improve the predictive value of the indices, it could result in significantly modified patterns of predictors and/or beta weights.

The predictive utility of the indices could be checked at several levels. An initial level would include scaling several prototype devices via the indices, collecting appropriate performance data (under conditions comparable to those employed in the original validation), and then measuring transfer performance on some intermediate device. The SQS 26CX and the SQS-4 might serve as prototypes with the SQS-23 as the transfer device. These would be particularly convenient and cost-effective since task-analytic data are already available (Wheaton and Mirabella, 1972). Similar procedures might also be employed with other surveillance devices such as ECM or radar which might, in fact, be preferable in order to test the generality of the predictive power of the indices.

Following such procedures, predicted and obtained performance scores would be compared. If the number of test devices were extended, then predicted and obtained performance scores could be compared correlatively.

Still a further level of corroborative analysis would include new estimates of beta weights based on a large sample (10 or more) of field devices. Each of these would have to be scaled and then subjected to performance tests. An alternative method would employ a smaller number of devices, reconfigured in a variety of ways in much the same manner that the synthetic sonar trainer was reconfigured to generate multiple tasks (e.g., by masking various controls and displays or by modifying the instructional sequences).

CONCLUSIONS AND IMPLICATIONS

The current research effort supported by the work of the preceding two phases provides a methodology for the predictive assessment of training device effectiveness. These efforts have demonstrated the feasibility of such a methodology by relating acquisition and transfer of procedural skills to variations in fourteen quantitative task indices. It has been possible to consistently obtain such relations using multiple regression techniques.

While the methodology is available immediately for limited use on the basis of laboratory validation, cross validation in the field remains to be and needs to be conducted. The discussion section has outlined a number of steps which can be taken in this direction. These include:

1. Applying the predictive methodology to several prototype trainers and contrasting actual with predicted performance scores.

2. Redetermining beta weights on a large sample of devices or a small number of devices which have been re-configured in the manner of the synthetic sonar trainer used in this research.

Even as the methodology is put into use, further validation and development would be of value. The thrust of such development might be to make the methodology applicable to other than procedural tasks.

In closing this discussion, a philosophical note should be sounded. The value of any tool for assessing training device effectiveness is constrained by the total system within which training takes place. The effectiveness of the predictions from the current methodology, for example, could be negated if selection procedures resulted in a particular range of student ability and that range were not taken into consideration. That is, the methodology emerging from this program deals with a small portion of the training systems problem. It is felt, however, that the portion covered is significant and important.

REFERENCES

Bernstein, B.R. and Gonzalez, B.K. Learning, retention and transfer in military training. Technical Report: NAVTRADEVCEN 69-C-0253-1, 1971. Naval Training Device Center, Orlando, Florida.

Dixon, W.J. (Ed.). Biomedical Computer Programs. Berkeley: University of California Press, 1968.

Farina, A.J.Jr., and Wheaton, G.R. Development of a taxonomy of human performance: The task characteristics approach to performance prediction. Technical Report No. 8, (1971). American Institutes for Research, Silver Spring, Maryland.

Fleishman, E.A., Teichner, W.H., & Stephenson, R.W. Development of a taxonomy of human performance: A review of the second year's progress. AIR-726-1/70-TPR2, 1970. American Institutes for Research, Silver Spring, Maryland.

Fowler, R.L., Williams, W.E., Fowler, M.G., & Young, D.D. An investigation of the relationship between operator performance and operator panel layout for continuous tasks. Report No. AMRL-TR-68-170, December, 1968. Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

Grimsley, D.L. Acquisition, retention, and retraining: Group studies on using low fidelity training devices. HumRRO Technical Report 69-4, March, 1969. Human Resources Research Office, Presidio of Monterey, California.

Mackie, R.R., and Harabedian, A. A study of simulation requirements for sonar operator trainers. Technical Report: NAVTRADEVCEN 1320-1, March, 1964. U.S. Naval Training Device Center, Orlando, Florida.

Mirabella, A. Two-dimensional versus three-dimensional tracking under two modes of error feedback. Journal of the Human Factors Society, (1969), 11, 9-12.

Folley, J.D.,Jr. Development of an improved method of task analysis and beginnings of a theory of training. Technical Report: NAVTRADEVCEN 1218-1, 1964. Naval Training Device Center, Port Washington, New York.

Prophet, W.W. and Boyd, H.A. Device-task fidelity and transfer of training: Aircraft cockpit procedures training. HumRRO Technical Report No. 70-10, July, 1970. Human Resources Research Organization, Fort Rucker, Alabama.

Siegel, A.I., and Federman, P.J. Validation of the DEI technique for large-scale display evaluation. Technical Report No. RADC-TR-67 134, May, 1967. Rome Air Development Center, Griffiss Air Force Base, New York.

Siegel, A.I., Miehle, W., & Federman, P. Information transfer in display-control systems, IV. Summary review of the DEI technique. (Fourth Quarterly Report). Applied Psychological Services, 1962a, Wayne, Pa.

REFERENCES (Cont'd)

Siegel, A.I., Miehle, W., & Federman, P. Information transfer in display-control systems, VI. A manual for the use and application of the display evaluative index. (Sixth Quarterly Report). Applied Psychological Services, 1962b. Wayne, Pa. AD 296374.

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

Wheaton, G.R., and Mirabella, A. Effects of task index variations on training effectiveness criteria. Technical Report: NAVTRADEVCEN 71-C-0059-1, 1972. Naval Training Device Center, Orlando, Florida.

Wheaton, G.R., Mirabella, A., & Farina, A.J., Jr. Trainee and instructor task quantification: Development of quantitative indices and a predictive methodology. Technical Report: NAVTRADEVCEN 69-C-0278-1, 1971. Naval Training Device Center, Orlando, Florida.

APPENDIX A

Task Characteristic Indices

TASK CHARACTERISTIC INDICES

1. MAIN* - defined as the number of responses comprising the main or dominant procedural sequence in an operations flow chart.

2. CNTG* - defined as the number of responses comprising the auxiliary or contingency procedural sequences.

3. TA - defined as the total number of responses (actions) comprising the procedural sequence in an operations flow chart. It represents the sum of MAIN and CNTG.

4. CONT - defined as the total number of different controls manipulated during performance of a subtask.

5. DISP - defined as the total number of different displays referenced during performance of a subtask.

6. E - defined as the total number of different equipment elements interacted with; this index is given by the sum of CONT and DISP.

7. LV - the link value reflecting the relative strength of the sequence of use among the various controls and displays. As used here, it is the sum of the products of the number of times a link is used, and the percentage of use of the link (Fowler, Williams, Fowler, & Young, 1968).

8. AA% - an index reflecting the percentage of alternative actions present in an operation. A score of "0% means that the highest number of alternative links are used, each with an equal frequency of use, and 100% score means there is only one link out of and into each control, with the same frequency used for all links" (Fowler et al., 1968).

9. F% - another index (Fowler et al., 1968) describing the extent to which all controls and displays are used an equal number of times (0%) or a theoretically defined optimum number of times (100%).

10. DEI - a measure of the effectiveness with which information flows from displays via the operator to corresponding controls. The index yields a dimensionless number representing a figure-of-merit for the total configuration of displays and controls (Siegel, Miehle, & Federman, 1962b).

11-13. D%, C%, E% - defined respectively as the number of display, control, or combined equipment elements which the operator actually employs relative to the total number of such elements which are available for use.

14-17. CRPS, FBR, INFO, INST* - refer to the frequency with which the operator makes various types of responses during performance of the task. Included are responses involving manipulation of controls (CRPS), securing of feedback (FBR), acquisition of information (INFO), as well as those primarily initiated by the instructor (INST).

---

* These indices were eliminated prior to analysis of Phase III data. Two of them, MAIN and CNTG correlated almost perfectly with TA and were eliminated for this reason. The third, INST, was invariant and eliminated for this reason.

APPENDIX B

Tasks Employed in Phases II and III

TASKS EMPLOYED IN PHASES II AND III

Three reference consoles provided the basis for the experimental tasks of the laboratory portions of Phases II and III. These were defined as the Complex (C) console, the Medium (M) console, and the Simple (S) console. Using these basic consoles, twenty trainee tasks were generated via a variety of manipulations. For example, indicator lights were retained in either: (1) all panels (a); (2) every second panel (s); (3) every third panel (t); or (4) none of the panels (n).

Tasks were also differentiated via different levels of embedding. For example, the simple task could be embedded either in the medium or complex console, while the medium task could be embedded only in the complex console.

Finally, any task based upon any of the above manipulations could be further reconfigured through the addition of special sequences of contingency actions.

Thus, a task based upon the simple console with indicator lights retained only on every third panel and with six additional contingency actions would be designated as Simple-third plus 6 or $S_t + 6$. If the same task were embedded in the complex console it would be designated as Simple-third plus 6 embedded in complex or $SE_{ct} + 6$.

LIST OF TASKS

1. Complex-all (Ca)
2. Complex-none (Cn)
3. Medium-all (Ma)
4. Medium-all embedded in complex (MEca)
5. Medium-third (M_t)
6. Medium-third plus 2 embedded in complex ($ME_{ct} + 2$)
7. Medium-none (M_n)
8. Medium-none embedded in complex ($ME_c n$)
9. Medium-none plus 2 ($M_n + 2$)
10. Simple-all (Sa)
11. Simple-all embedded in medium (SEma)
12. Simple-all embedded in complex (SEca)
13. Simple-second (S_s)
14. Simple-second embedded in medium (SEms)
15. Simple-second embedded in complex (SEcs)
16. Simple-third plus 6 ($S_t + 6$)
17. Simple-third plus 6 embedded in complex ($SE_{ct} + 6$)
18. Simple-none (Sn)
19. Simple-none embedded in medium (SEmn)
20. Simple-none embedded in complex (SEcn)

APPENDIX C

Data Arrangements Employed in the Training Methods Study

DATA ARRANGEMENTS EMPLOYED IN THE TRAINING METHODS STUDY*

Analysis A

Method	Task Complexity Level	
	Simple (Sa)	Complex (Ca)
Pictorial Cold Panel Hot Panel		

Analysis B

Method	Task Complexity Level			
	Simple (S)		Medium (M)	
	No Embedding (Sa)	Embedding (SEca)	No Embedding (Ma)	Embedding (MEca)
Pictorial Cold Panel Hot Panel				

Analysis C

Method	Level of Embedding		
	No Embedding (Sa)	Moderate Embedding (SEma)	High Embedding (SEca)
Pictorial Cold Panel Hot Panel			

---

* Note that these matrices are not entirely independent since some experimental groups are used more than once.

APPENDIX D

Application of the Methodology

APPLICATION OF THE METHODOLOGY

The purpose of this appendix is to outline the procedures required to apply the 17-index battery developed by the project and to define some constraints on its use.

First, it should be emphasized that the battery is most applicable to procedural tasks. Results of the field studies indicate that some tasks such as target recognition are not well differentiated on the basis of these particular indices.

Second, it should be noted that a figure-of-merit approach, in the most literal sense, is not appropriate. Our research showed, at least for the limited set of devices looked at, that sub-tasks must be defined for the device to be quantified. The indices are then applied to the sub-tasks rather than to the device as a whole. Thus device evaluation may require multiple judgments, or at least a sub-task specific judgment.

Third, multiple criteria of device effectiveness are potentially available. A choice among these is necessary since the pattern of predictors may change from criterion to criterion. In particular, the different criteria include measures of speed and accuracy at various stages of training and transfer.

PROCEDURES FOR DEVICE QUANTIFICATION

STEP 1: TASK DEFINITION. Define the tasks or sub-tasks associated with the device: These usually will consist of conventionally recognized sets of operations. The distinctions among the sets often will be made arbitrarily, but unavoidably in order to carry out task analysis. Thus, for surveillance trainers, sub-tasks would include set-up, detection, localization, and classification. For flight trainers, the sub-tasks might include set-up (check-out), take-off, landing, emergency procedures, and navigation. The quantification procedures require that the sub-tasks be viewed as independent, even though in an operational sense they overlap or interact.

STEP 2: DATA COLLECTION. Data collection consists of completing the appended Task Analysis Data Form (Appendix D-1) for each sub-task to be examined. Identification information is entered at the top of the form, and in the table below, each sequential response in the sub-task is listed and described.

The data collector begins his operation by labeling each display and control on the panel under consideration. Where distinctive parts of a given display or control are identifiable each part is given a separate number. For example, on a time-bearing paper recorder, equipped with a bearing rate indicator, the T-B chart and the B-R indicator are labeled separately. *

A qualified instructor then proceeds to describe the specific sub-task, after being provided with the appropriate instructional set. That is, he must view the sub-task as independent of other sub-tasks and he must sequentially name and describe each response. In each statement, the instructor should name the equipment element, its assigned number, the action involved, the number of states which the display or control can assume, and the number of states which the trainee is normally called upon to deal with. Where contingency actions follow, each contingency should be described in the same amount of detail, as indicated above.

For example, the instructor might say:

> "Turn No. 1, the on-off switch to the ON position".
>
> Check No. 2, the POWER ON indicator for a red indication.
>
> Read No. 3, the POWER LEVEL METER for voltage level. Meter is calibrated in .10 volt units. Meter is normally read in .50 volt units. Voltage range is 0 to 10 volts.
>
> If meter exceeds 5 volts, turn No. 1, the ON-OFF switch to the OFF position and request maintenance. Otherwise, proceed to next action.

These statements would be summarized by the data collector as shown in the appended Task Analysis Data Form (Appendix D-1).

---

* The data collector will generate a list of all displays and controls. For each equipment element in the list the following data should be recorded.

a. The labeled code number of the control or display involved in the response action (i.e., 1, 2, 3, etc.).

b. Designation of the equipment as a control, a display, or a combination of both (i.e., C. D. B.).

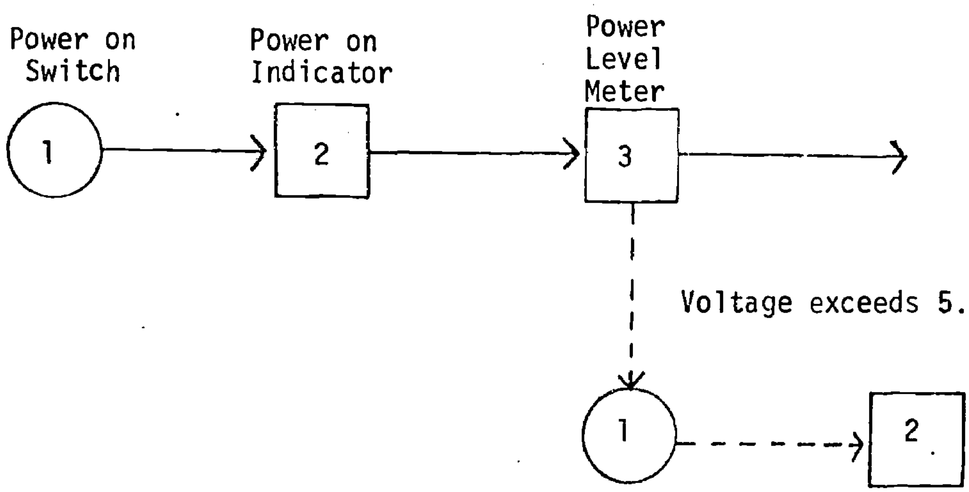
c. The nomenclature of the equipment involved (i.e., sea-state noise level filter).

d. The type of hardware which the equipment represents together with the states it can assume (i.e., a ten position rotary knob - 1, 2, . . ., 9, 10).

This listing can be facilitated by a form similar to the one shown in Appendix D-2.

STEP 3: DATA FLOW CHARTING. The information provided by Appendix D-1 can be collected in the alternative form of a flow chart. This form is particularly useful as an aid in generating the indices of Fowler, et al. (1968).

The flow chart consists of a linear sequence of circles and squares representing main line actions or responses. Squares represent display readings or judgments while circles represent manipulations of controls. Contingency actions are shown by squares and/or circles displaced below the main line of action, and connected by dotted lines. Thus the data in Appendices D-1 and D-2 would be represented as follows.



Additional detail on this procedure is provided in Wheaton, Mirabella and Farina (1971).

STEP 4: COMPUTATION OF DISPLAY EVALUATIVE INDEX (DEI). The amount of detail and complexity involved in computing the index are too extensive for presentation here. The reader is therefore referred to the manual authored by Siegel, Miehle, and Federman (1962b). The manual contains step-by-step instructions for applying DEI, plus computational examples and a glossary. Additional information is provided by Wheaton, Mirabella, and Farina (1971). However, the steps in this application will be outlined here.

DEI is a method for measuring the effectiveness with which information is transmitted between an operator and his console. It is a dimensionless index varying from 0 to 1. In general, the technique requires that displays be represented symbolically in one column, controls in an adjacent column, and a variety of links drawn between the displays and controls. These links are then quantified and tabulated in a variety of ways to arrive ultimately at a single value. The initial representation of displays and controls is in the form of a Transfer Chart (Appendix D-3). Here displays are shown by circles on the left, controls are shown by triangles on the right, with intervening operations represented between them. These operations include computations, comparisons among displays, combinations of display readings and table look-up operations. Links are drawn from display symbols to intervening symbols, and from intervening symbols to control symbols. Links are also drawn directly between displays and controls.

Quantification proceeds with the aid of a link table (illustrated in Appendix D-4). Here the links are listed and quantified in a variety of ways. These include display and/or control resolution which is the $\log_2 n$, where n is the number of states that a display or control can assume. This value is calculated for each display and control. Any discrepancy between these values for a given link is listed in the mismatch column. Next a link weight is assigned, depending upon the type of link involved. Definitions of the different link types and their weights are given in Siegel, et al. (1962b).

Finally, a DEI worksheet (illustrated in Appendix D-5) is prepared. The computations listed in this worksheet are based upon information in the transfer table.

STEP 5: COMPUTATION OF PANEL LAYOUT INDICES. Details and illustrations of this procedure are presented in Fowler, et al. (1968) and in Wheaton, Mirabella and Farina (1971).

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. 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 flow chart information into a Link Value Table (Appendix D-6). Each link in the flow chart is listed in coded form in column 1 of the Link Value Table. The first number in the code refers to the display or control from which a given link <u>leaves</u>. The second number refers to the hardware component which the link then <u>enters</u>. In columns 2, 3, and 4 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 <u>leaving</u> a given control or display; and (3) a link value which is the product of data recorded in the second and third columns. In columns 5, 6, 7, and 8, check marks are entered to indicate whether each link value is: (1) the maximum value leaving a control and entering a display; (2) the maximum value entering; (3) the maximum value leaving; or (4) none of the cases above.

The information in the link table 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. (1968). 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. From this layout and the link table, it is possible to compute LV, AA% and F%.

STEP 6: DERIVATIVE INDICES. The indices of Siegel and Fowler represent four of those in the battery: DEI, AA%, F%, LV. The remaining 13 indices are derivatives of the methodology involved in the first two cases. They are obtained in the following manner.

Total Actions (TA) equal the sum of all links defined by the Fowler link chart. These consist of primary (MAIN) and contingency (CNTG) responses.

Numbers of controls (C), displays (D), and their combination (E) are obtained by counting circles and squares in the Fowler panel lay-out chart. The total numbers of displays and controls for the (D%), (C%), and (E%) indices are proportional values based upon those used relative to those available on the operator panel under consideration.

Number of Control Responses (CRPS) equals the number of links entering circles on the sub-task flow chart.

Number of feedback responses (FBR), number of information acquisition responses (INFO) and number of instructor initialed responses are obtained from the Task Analysis Data Form (Appendix D-2).

APPENDIX D-1

TASK ANALYSIS DATA FORM

Date	6/1/73	Training Device #	6 x 3	Trainee Task	
Instructor	Jones, CPO	Training Device Name	Bravo	SU ✓ D ___ L ___ C ___	
Data Recorder	Mitchell	Operational Equip. #		Page No.	1
		Location	NTEC		

Resp. #	Response Description (action, equip., equip. #, alternative outcomes	Ctg. Resp. #	Contingency Response Description (action, equip., equip. #, alternative outcomes
1	#1 to ON		
2	#2 - check for red		
3	#3 - read to .50 volt (0-10 volts in	3.1.a.	If #3 $\geq$ 5 volts, #1 to OFF
	.10 units).	3.1.b.	Check for OFF indication on indicator #2

APPENDIX D-2

TASK ANALYSIS DATA FORM - EQUIPMENT ELEMENTS

Date ________ Training Device # ________________ Trainee Task,

Instructor ________ Training Device Name ________________ SU___ D___ L___ C___

Data Recorder ________ Operational Equip. # ________________ Page No. ________

Location ________________

Rep. #	Equip #	C,D	Equip. Nomenclature	Hardware & Units	Function	Cont-rol Resp.?	Feed-Back ?	Info Acq. ?	Instruc. Initia-ted?	Comments
	1	C	Power On Switch	2 Position Toggle	Energize Console					
	2	D	Power On Indicator	2 State Pilot Light	Indicate Status of #1					
	3	D	Power Level Indicator	Meter, 0-10 volts in .10 units	Indicates voltage to the stock, monitored for safe level					

APPENDIX D-3

TRANSFER CHART FOR DEI

Device ______________________ Date ______________________

Sub-Task ______________________ Location ______________________

Displays	Intervening Processes	Controls

APPENDIX D-4

LINK TABLE FOR DEI

Device Number ____________

Sub-Task ________________ Page Number ________

Link No.	Link Type	Display Info		Control Info		MisMatch	Link Weight
		No. States	No. Digits	No. States	No. Digits	(|Mi|)	(Wi)
N=					$2\|Mi\| =$		$\sum Wi=$

APPENDIX D-5

DEI WORKSHEET

Device Number ______________________ Date ______________________

Sub-Task ______________________ Page No. ______________________

$N_1 = (n + m)\ u =$ . . . . ______________ (N_1)

Sum $|mi|$ . . . . ______________

1/4 Sum $|mi|$ . . . . ______________

$N_2 = \exp(-\ 1/4\ \text{Sum}\ |mi|)$ . . . . ______________ (N_2)

$N_3 = (1 + wi)$ . . . . ______________ (N_3)

$N_4 = (N)$ . . . . ______________ (N_4)

$N_5 = (n + m)\ t$ . . . . ______________ (N_5)

$(Q) =$ ______________

$(n_o) =$ ______________

$N_6 = (Q + n_o)$ . . . . ______________ (N_6)

$$DEI = \frac{N_1 \cdot N_2}{N_3 \sqrt{N_4 \cdot N_5 \cdot N_6}}$$

APPENDIX D-6

LINK VALUE TABLE

1 Links	2 No. Times Link Used	3 % Use	4 Link Value	5 Max. Link Value In & Out	6 Max. Link Value In Only	7 Max. Link Value Out Only	8 Remainder
1-2	2	100	200				
2-3	1	100	100				
3-1	1	100	100				

APPENDIX E

Multiple Regression Equations

APPENDIX E-1

PREDICTION OF RESIDUAL ACQUISITION TIME AND ERROR SCORES FOR FIRST, MIDDLE, AND LAST BLOCK OF ACQUISITION TRIALS

Trial Block	Predicted Score	=	Constant	+	1st Predictor	+	2nd Predictor	+	3rd Predictor
			Time Scores						
T_{1-2}		=	249.49789	-	0.79367 (E%)	-	5.88988 (DEI)	-	4.28367 (CONT)
T_{7-8}		=	262.60815	-	0.80009 (C%)	-	1.99320 (F%)	-	2.24697 (INFO)
T_{13-15}		=	69.51535	-	0.36651 (C%)	-	1.78827 (DEI)	-	1.20765 (DISP)
			Error Scores						
T_{1-2}		=	20.14337	-	0.27424 (E%)	-	0.10844 (F%)	+	0.14615 (D%)
T_{7-8}		=	3.98721	-	0.15948 (DEI)	-	0.11061 (FBR)	-	0.00942 (C%)
T_{13-15}		=	7.81576	-	0.05932 (DEI)	-	0.08895 (DISP)	-	0.08643 (AA%)

APPENDIX E-2

PREDICTION OF RAW ACQUISITION TIME AND ERROR SCORES FOR FIRST, MIDDLE AND LAST BLOCK OF ACQUISITION TRIALS

Trial Block	Predicted Score	=	Constant	+	1st Predictor	+	2nd Predictor	+	3rd Predictor
					Time Scores				
T_{1-2}		=	358.47632	-	8.64873 (DEI)	+	3.06072 (FBR)	-	0.53596 (E%)
T_{7-8}		=	123.40608	-	4.18158 (DEI)	+	2.53731 (E)	-	0.36898 (C%)
T_{13-15}		=	77.40903	+	1.23966 (TA)	-	2.05052 (DEI)	-	0.33148 (C%)
					Error Scores				
T_{1-2}		=	10.74920	-	0.34075 (DEI)	+	0.00108 (LV)	-	0.03568 (E%)
T_{7-8}		=	2.79676	-	0.17705 (DEI)	+	0.04663 (CRPS)	-	0.04156 (FBR)
T_{13-15}		=	7.13149	+	0.03994 (CRPS)	-	0.10182 (AA%)	-	0.04808 (DEI)

APPENDIX E-3

PREDICTION OF RAW TRANSFER TIME AND ERROR SCORES FOR FIRST, MIDDLE, AND LAST BLOCKS OF TRANSFER TRIALS USING DIFFERENCES BETWEEN ACQUISITION AND TRANSFER TASK INDICES AS PREDICTORS AND WEIGHTING THE DEI INDEX

Trial Block	Predicted Score	=	Constant	+	1st Predictor	+	2nd Predictor	+	3rd Predictor
					Time Scores				
T_{1-2}		=	122.88995	+	0.46327 (DISP)	+	1.01835 (C%)	+	3.79809 (DEIW)
T_{5-6}		=	108.42064	+	0.92285 (DISP)	+	0.50996 (C%)	+	1.80256 (DEIW)
T_{9-10}		=	102.63283	+	2.23652 (DISP)	+	0.82897 (C%)	-	0.36180 (D%)
					Error Scores				
T_{1-2}		=	1.84047	+	0.42065 (DEIW)	-	0.40431 (INFO)	+	0.13797 (FBR)
T_{5-6}		=	1.12822	+	0.24981 (DEIW)	-	0.12846 (E)	+	0.05339 (F%)
T_{9-10}		=	0.94322	+	0.19687 (DEIW)	-	0.11582 (E)	+	0.01126 (D%)

APPENDIX E-4

PREDICTION OF RAW TRANSFER TIME AND ERROR SCORES FOR FIRST, MIDDLE, AND LAST BLOCKS OF TRANSFER TRIALS USING DIFFERENCES BETWEEN ACQUISITION AND TRANSFER TASK INDICES AS PREDICTORS DEI UNWEIGHTED

Trial Block	Predicted Score	=	Constant	+	1st Predictor	+	2nd Predictor	+	3rd Predictor
					Time Scores				
T_{1-2}		=	125.03101	+	0.17039 (DISP)	+	0.71444 (C%)	+	2.80941 (FBR)
T_{5-6}		=	110.91982	+	1.25452 (DISP)	+	0.36835 (C%)	+	1.56406 (DEI)
T_{9-10}		=	102.63283	+	2.23652 (DISP)	+	0.82897 (C%)	-	0.36180 (D%)
					Error Scores				
T_{1-2}		=	2.31813	+	0.49029 (DEI)	-	0.44035 (E)	+	0.43373 (DISP)
T_{5-6}		=	1.29935	+	0.32803 (DEI)	-	0.29532 (E)	+	0.24938 (DISP)
T_{9-10}		=	1.15513	+	0.18036 (DEI)	-	0.21853 (E)	+	0.19267 (DISP)

APPENDIX E-5

PREDICTION OF RAW TRANSFER TIME AND ERROR SCORES FOR FIRST, MIDDLE AND LAST BLOCK OF TRANSFER TRIALS USING ABSOLUTE VALUES ON ACQUISITION TASKS AS PREDICTORS

Trial Block	Predicted Score	=	Constant	+	1st Predictor	+	2nd Predictor	+	3rd Predictor
			Time Scores						
T_{1-2}		=	718.02734	-	24.50572 (E)	+	34.71390 (INFO)	-	2.64899 (F%)
T_{5-6}		=	389.26636	-	11.24068 (E)	+	15.40897 (INFO)	-	1.13062 (F%)
T_{9-10}		=	208.42519	-	6.02910 (E)	+	2.34210 (TA)	-	0.20883 (C%)
			Error Scores						
T_{1-2}		=	27.70270	-	0.29680 (FBR)	+	0.03198 (D%)	-	0.33444 (AA%)
T_{5-6}		=	0.37502	-	0.19953 (FBR)	+	0.02062 (D%)	+	0.12596 (INFO)
T_{9-10}		=	0.05431	-	0.12145 (FBR)	+	0.09846 (INFO)	+	0.01112 (E%)

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) American Institutes for Research Silver Spring, Maryland	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

EFFECTS OF TASK INDEX VARIATIONS ON TRANSFER OF TRAINING CRITERIA

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Report

5. AUTHOR(S) (First name, middle initial, last name)

Angelo Mirabella, George R. Wheaton

6. REPORT DATE January 1974	7a. TOTAL NO. OF PAGES 83	7b. NO. OF REFS 16

8a. CONTRACT OR GRANT NO. N61339-72-C-0126 b. PROJECT NO. 1752-03 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

Approved for public release; distribution is unlimited.

11. SUPPLEMENTARY NOTES There are three reports in this series. The other reports are NAVTRAEQUIPCEN 71-C-0059-1 and NAVTRADEVCEN 69-C-0278-1.	12. SPONSORING MILITARY ACTIVITY Naval Training Equipment Center Orlando, Florida

13. ABSTRACT

This report describes the concluding study in a three phase program. The goal of the program has been to develop and validate a set of quantitative task indices for use in forecasting the effectiveness of training devices.

In Phase I, indices were collated and applied to an assortment of passive - and active-sonar training devices. On the basis of these field applications, 17 measures were chosen. In Phase II, training of procedural skills in a modularized, synthetic sonar trainer was studied in the validation of the 17-index battery. Substantial and significant multiple correlation coefficients were obtained for both performance time and errors during skill acquisition.

Phase III - the index battery was validated against transfer of training criteria. Phase III results demonstrated that quantitative variations in task design could be related significantly and substantially to variations in transfer of training measures. On the basis of these results and those of Phase II, a set of predictive equations was constructed.

It was concluded that these equations could be employed immediately to compare the efficacy of competing trainer prototypes, but that additional validation efforts in the field were necessary in order to extend confidence and generality of the methodology. It was further concluded that the battery could be useful in selecting tasks for research on the interaction of task variables and other training system variables. A demonstration of this application was carried out in which training method was studied as a function of task complexity. Results of this latter study provided some support for the hypothesis that the effectiveness of dynamic versus static procedural training varied with changes in task parameters.

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Task Quantification						
Training Device Effectiveness						
Transfer of Training						

HUMAN FACTORS DISTRIBUTION LIST

NOTE

Mailing labels are prepared, when needed, as a computer listing, the source of which is updated on a weekly basis. It is not practical to prepare distribution lists each time labels are prepared. Therefore such lists are prepared semiannually, and a slight discrepancy may exist between the addressees on this list and those appearing on the labels used to distribute this publication.

Chief of Naval Training
Attn: Captain A. E. McMichael, N-3
Pensacola, FL 32508

Chief of Naval Training
Attn: Captain B. C. Stone
Pensacola, FL 32508

Chief of Naval Training
Attn: Dr. W. Maloy, Code 01A
Pensacola, FL 32508

Chief of Naval Material
Attn: Mr. A. L. Rubinstein, MAT-03424
Navy Department
Washington DC 20360

Commanding Officer
Naval Submarine Base, New London
Attn: Psychology Section
Box 00
Groton, CT 06340

Chief of Naval Air Training
Attn: Joseph L. Ulatoski
Naval Air Station
Corpus Christi, TX 78419

Commander Training Command
Attn: Educational Advisor
U.S. Pacific Fleet
San Diego, CA 92147

Chief Naval Research
Attn: Code 458 Navy Department
Arlington, VA 22217

COMDT of Marine Corps
Code A03C
Washington, DC 20380

USAF Human Res Lab
AFHRL DO Operations Ofc
Brooks AFB, TX 78235

CDR NAVAIR DEVEL CTR
Attn: Human Engrg Br
Warminister, PA 18974

Commander
Naval Electronics Systems Command
Code 03
Washington, DC 20360

NAVAERSPMEDINST
NAVAEROSPREGMEDCEN
ATTN: Ch Avia Psych Div
Pensacola, FL 32512

Chief Naval Operations
Attn: CDR H. J. Connery
OP-987 M42 Navy Dept
Washington, DC 20350

HUMAN FACTORS DISTRIBUTION LIST

Commander Training Command
Attn: Educational Advisor
U. S. Atlantic Fleet
Norfolk, VA 23511

Assistant Secretary of the Navy (R&D)
Attn: Dr. Samuel Koslov, 4E741
Navy Department
Washington, DC 20350

Director of Defense Research & Engineering
Attn: Lt Col Henry Taylor, OAD (R&D)
Washington, DC 20301

Dir Human Res Rsch Org
300 N Washington St
Alexandria, VA 22314

Commander
NAVAIRSYSCOM
Code 03
Washington, DC 20360

CDR NAVSHIPS SYS CMD
NAVSHIPS SYS CMD HQS
SHIPS 03H Geo. N. Graine
Washington, DC 20360

Chief Naval Operations
Attn: Dr. J. J. Collins
OP-987F Navy Dept
Washington, DC 20350

ERIC Clearinghouse UN
EDUCAT MEDIA--TECH
Stanford Univ
Stanford, CA 94305

Bureau Naval Personnel
Attn: PERS A3
Arlington Annex
Washington, DC 20370

Chief Naval Research
Psychological Sciences
Code 450 Navy Dept
Arlington, VA 22217

Chief Naval Material
MAT 031M
Washington, DC 20360

USAF Human Res Lab
Personnel Rsch Div
Lackland AFB, TX 78236

Human Resources Resch
Orgztn Div 6 Aviation
P. O. Box 428
Fort Rucker, AL 36360

National Science Foundation
Attn: Dr. Henry S. Odbert
1800 G St NW
Washington, DC 20550

Chief Naval Operations
Attn: M. K. Malchorn
OP-14C Navy Dept
Washington, DC 20350

USAF Human Res Lab
AFHRL-FT
Flying Trng Div
William AFB, AZ 85224

USAF Human Res Lab
AFHRL-TT
Tech Tng Div
Lowry AFB, CO 80230

Commander
NAVORDSYSCMD
Code 03
Washington, DC 20360

HUMAN FACTORS DISTRIBUTION LIST

HQS AF SYS CMD DLSL
Ofc Scientific Rsch
Andrews AFB
Washington, DC 20331

Commander
NAVSUPSYSCMD
Code 03
Washington, DC 20360

CH RSCH AND DEVLPMT
CH Behavioral Sci Div
Dept of Army
Washington, DC 20310

Human Res Rsch Org
Div No. 1 Sys Oper
200 N. Washington ST
Alexandria, VA 22314

CO NAV MISSILE CTR
Attn: Hd Human Factors
Engrg Br
Point Mugu, CA 93042

Commanding Officer
Navy Medical
Neuropsychiatric Resch Unit
San Diego, CA 92152

CO NAVAIR TECH TRNG
NAS Memphis
Attn: Dr. G. D. Mayo
Hd Research Br
Millington, TN 38054

Dir DEF RSCH-ENGRG
ARPA
Behavioral Science Div
Attn: LCOL A. W. Kibler
Washington, DC 20301

Scientific Technical
Information Office
NASA
Washington, DC 20546

Behavioral Sci Div
Ofc Chief Rsch - Devel
Dept of Army
Washington, DC 20310

Chief of Naval Technical Training
NAS Memphis 75
Attn: Code 34 Dr. Harding
Millington, TN 38054

Dr. John Meyer
HDQTRS Air Training Command
XPT
Randolph AFB, TX 78148

Joseph J. Cowan
Ch PERS RSCH BRANCH
USCG HQ PO-1 STA 3-12
400 Seventh St. S. W.
Washington, DC 20590

Executive Editor
Psycho Abstracts
American Psych Assoc
1200 17th St NW
Washington, DC 20036

Dr. Ralph R. Canter
Dir MIL MANPWR RSCH
OSAD M-RA MR-U
Pentagon RM 3D960
Washington, DC 20301

Dr. John W. Weisz
Dir Human Engrg Lab
USA Aberdeen Rsch Devel Ctr
Aberdeen Proving Grounds
MD 21005

HUMAN FACTORS DISTRIBUTION LIST

USAF Human Res Lab
AFHRL-TR
Trng Res Div
Wright-Patterson AFB, OH 45433

U.S. Army Research Institute
Research Lab
Commonwealth Bldg (Rm 239)
1320 Wilson Blvd
Arlington, VA 22209

Commanding Officer
PERS-TRNG RESH DEV LAB
San Diego, CA 92152

NAV PERSONNEL RESCH AND
DEVELOPMENT LABORATORY
Attn: Library
Bldg 200 Rm 3307 WNY
Washington, DC 20390

CH NAVTRASUP
Code N-2, Bldg 45
(Dr. Charles Havens)
NAS Pensacola, FL 32508

GENERAL DISTRIBUTION

Naval Training Equipment Center 63
Orlando, Florida 32813

Defense Documentation Center 12
Cameron Station
Alexandria, Virginia 22314

Naval Training Equipment Center, Orlando, Fla.

TR NAVTRAEQUIPCEN 72-C-0126-1 UNCLASSIFIED
AD___________

EFFECTS OF TASK INDEX VARIATIONS ON TRANSFER OF TRAINING CRITERIA. FINAL REPORT.
83p, 5 tables, 19 illus., 16 refs.

This report describes the concluding study in a three phase program. The goal of the program has been to develop and validate a set of quantitative task indices for use in forecasting the effectiveness of training devices.
In Phase I the indices were defined and in Phase II substantial and significant multiple correlation coefficients were obtained between task indices and both performance time and errors during skill acquisition.
Phase III - the index battery was validated against transfer of training criteria.

KEY WORDS

Task quantification
Training device effectiveness
Transfer of training

American Institutes for Research
Mirabella, Angelo
Wheaton, George R.
N61339-72-C-0126
NAVTRAEQUIPCEN
Task No. 1752-03

Results demonstrated that quantitative variations in task design could be related significantly and substantially to variations in transfer of training measures. On the basis of these results and those of Phase II, a set of predictive equations was constructed.
It was concluded that these equations could be employed immediately to compare the efficacy of competing trainer prototypes but that additional validation efforts in the field were necessary in order to extend confidence and generality of the methodology. It was further concluded that the battery could be useful in selecting tasks for research on the interaction of task variables and other training system variables. A demonstration of this application was carried out in which training method was studied as a function of task complexity. Results of this latter study provided some support for the hypothesis that the effectiveness of dynamic versus static procedural training varied with changes in task parameters.

Naval Training Equipment Center, Orlando, Fla.

TR NAVTRAEQUIPCEN 72-C-0126-1 UNCLASSIFIED
AD____________

EFFECTS OF TASK INDEX VARIATIONS ON TRANSFER OF TRAINING CRITERIA. FINAL REPORT. 1974
83p, 5 tables, 19 illus., 16 refs.

This report describes the concluding study in a three phase program. The goal of the program has been to develop and validate a set of quantitative task indices for use in forecasting the effectiveness of training devices.
In Phase I the indices were defined and in Phase II substantial and significant multiple correlation coefficients were obtained between task indices and both performance time and errors during skill acquisition.
Phase III - the index battery was validated against transfer of training criteria.

Results demonstrated that quantitative variations in task design could be related significantly and substantially to variations in transfer of training measures. On the basis of these results and those of Phase II, a set of predictive equations was constructed.
It was concluded that these equations could be employed immediately to compare the efficacy of competing trainer prototypes but that additional validation efforts in the field were necessary in order to extend confidence and gen[illegible]lity of the methodology. It was further conclude[illegible] that the battery could be useful in selecting tasks for research on the interaction of task variables and other training system variables. A demonstration of this application was carried out in which training method was studied as a function of task complexity. Results of this latter study provided some support for the hypothesis that the effectiveness of dynamic versus static procedural training varied with changes in task parameters.

KEY WORD

Task quanti
Training device effectiveness
Transfer of training

American Institutes for Research
Mirabella, Angelo
Wheaton, George R.
N61339-72-C-0126
NAVTRAEQUIPCEN
Task No. 1752-03

BEST COPY AVAILABLE

Naval Training Equipment Center, Orlando, Fla.

TR NAVTRAEQUIPCEN 72-C-0126-1 UNCLASSIFIED
AD__________

EFFECTS OF TASK INDEX VARIATIONS ON TRANSFER OF TRAINING CRITERIA. FINAL REPORT.
83p, 5 tables, 19 illus., 16 refs.

This report describes the concluding study in a three phase program. The goal of the program has been to develop and validate a set of quantitative task indices for use in forecasting the effectiveness of training devices.
In Phase I the indices were defined and in Phase II substantial and significant multiple correlation coefficients were obtained between task indices and both performance time and errors during skill acquisition.
Phase III - the index battery was validated against transfer of training criteria.

KEY WORDS

Task quantification
Training device effectiveness
Transfer of training

American Institutes for Research
Mirabella, Angelo
Wheaton, George R.
N61339-72-C-0126
NAVTRAEQUIPCEN
Task No. 1752-03

Results demonstrated that quantitative variations in task design could be related significantly and substantially to variations in transfer of training measures. On the basis of these results and those of Phase II, a set of predictive equations was constructed.
It was concluded that these equations could be employed immediately to compare the efficacy of competing trainer prototypes but that additional validation efforts in the field were necessary in order to extend confidence and generality of the methodology. It was further concluded that the battery could be useful in selecting tasks for research on the interaction of task variables and other training system variables. A demonstration of this application was carried out in which training method was studied as a function of task complexity. Results of this latter study provided some support for the hypothesis that the effectiveness of dynamic versus static procedural training varied with changes in task parameters.

BEST COPY AVAILABLE