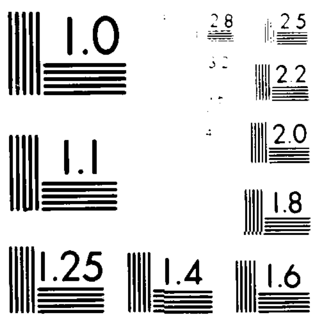
AD-A108 499 CANYON RESEARCH GROUP INC WESTLAKE VILLAGE CA F/G 5/9
APPLICATION OF A MULTIFACTOR APPROACH TO TRANSFER OF TRAINING R--ETC(U)
JUL 81 C W SIMON, S N ROSCOE N61339-78-C-0060
UNCLASSIFIED CWS-01-82 NAVTRAEQUIPC-78-C-0060-6 NL

1 OF 1
AD A

END
DATE
FILMED
01-82
DTIC



MICROCOPY RESOLUTION TEST CHART

(12) LEVEL II

YW

AD A108499

Technical Report NAVTRAEQUIPCEN 78-C-0060-6

APPLICATION OF A MULTIFACTOR APPROACH TO TRANSFER OF TRAINING RESEARCH

Charles W. Simon
Canyon Research Group, Inc.
741 Lakefield Road, Suite B
Westlake Village, California 91361

and

Stanley N. Roscoe
ILLIANA Aviation Sciences, Ltd.
2090 Wyoming Avenue
Las Cruces, New Mexico 88001

TECHNICAL REPORT MAY 1980 - July 1981

July 1981

DoD Distribution Statement

Approved for public release;
distribution unlimited.

NAVAL TRAINING EQUIPMENT CENTER
ORLANDO FLORIDA 32813

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NAVTRAEQUIPCEN 78-C-0060-6	2. GOVT ACCESSION NO. AD-A108499	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) APPLICATION OF A MULTIFACTOR APPROACH TO TRANSFER OF TRAINING RESEARCH		5. TYPE OF REPORT & PERIOD COVERED Technical Report 1 May 1980 - 31 July 1981
		6. PERFORMING ORG. REPORT NUMBER CWS 01-82
7. AUTHOR(s) Charles W. Simon and Stanley N. Roscoe		8. CONTRACT OR GRANT NUMBER(s) N 61339-78-C-0060
9. PERFORMING ORGANIZATION NAME AND ADDRESS Canyon Research Group, Inc. 741 Lakefield Road, Suite B Westlake Village, CA 91361		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NAVTRAEQUIPCEN Project No. 4781-6P1A
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Training Equipment Center Orlando, Florida 32813		12. REPORT DATE July 1981
		13. NUMBER OF PAGES 83
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Training and transfer data relationships
Economical multifactor, multicriterion study
Visual Technology Research Simulator (VTRS)
Transfer of training
Experimental designs
Research methods

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Multifactor transfer of training experiments are expensive to perform because of the large number of subjects required, the extended training they receive, and their subsequent operational testing. More economical data collection techniques are needed to fulfill the mission of the Naval Training Equipment Center's Visual Technology Research Simulator.

(over)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Continued

An experimental effort was undertaken: (a) to establish relationships among training, test, and transfer scores in the context of the manual control of a maneuvering vehicle; (b) to determine the relative complexities of response surfaces for training, test, and transfer; (c) to demonstrate a new transfer research paradigm that makes economically feasible the simultaneous investigation of the effects of a large number of equipment design variables on the transfer to multiple test configurations; and (d) to extract from the available data indications that will enable the transfer effectiveness of simulator characteristics to be estimated with minimum costs.

A horizontal tracking task was used in the study. Six factors were varied to form 49 simulator training configurations. These factors included five dynamic simulator design variables: vehicle control order, display lag, tracking mode (pursuit vs. compensatory), prediction time, and control gain. The sixth variable was the number of training trials given before transition to one of three transfer vehicle configurations, designated Hard, Central, and Easy.

Eighty flight-naive adult males participated in the experiment. Each of 48 was trained and tested on a different combination of training and transfer simulator configurations. Eight more were trained and tested on the same 49th combination. Data collected at these 49 points provided estimates of all main effects and two-factor interactions for both the training and transfer configurations. Additional data were collected from three control groups who received no prior training.

In this initial experiment the following relationships between training and transfer performance were indicated: (a) the transfer surface appeared less complex that the training surface; (b) the relationship between training and transfer scores was positive but too weak for predictive purposes, (c) some factors had large effects in training and small effects in transfer, and vice versa, and (d) transfer was facilitated when the values of certain variables result in training conditions that were more difficult than the subsequent transfer criterion conditions. Future investigations will help determine their generality.

The study demonstrated the efficiency and economy in collecting multifactor, multicriterion transfer of training data. This type of experiment is particularly useful in the early stages of a simulator design program when many alternatives should be considered and the individual contributions of component design parameters should be evaluated separately from overall simulator effectiveness.

FOREWORD

One of the great drawbacks associated with training device and training technology research is the difficulty of conducting transfer of training experiments. The reasons are well-known: for instance, research that involves training people to use operational equipment is costly, complicated and logistically difficult; the use of a large number of subject participants is usually necessary; and very few experimental factors can be examined at one time. The result: a great deal of money is spent to do relatively few experiments, and our progress in understanding issues of trainer design and usage is slow.

This report is concerned with two means by which progress may be accelerated: (1) using more efficient experimental designs in order to examine more factors simultaneously; and (2) understandng the relationships between performance and transfer data for different classes of variables, in order to predict transfer on the basis of performance data. The first author, Dr. Simon, has worked for more than a decade to study and publicize methods for conducting more informative and efficient performance experiments. The second author, Dr. Roscoe, has devoted most of his professional life to aviation training research. They have collaborated to produce the first reported attempt to use an economical multifactor design to explore issues of training and transfer.

Psychologists involved in applications-oriented simulation and training research will find this report thought provoking, both in terms of metholodogy and results. Considering the number of variables examined, it was the most complex transfer of training experiment ever done, and serves to show that it is no longer necessary to think solely in terms of small-factor studies -- there are alternatives. The results provide some information on a variety of performance-transfer relationships (e.g., relative complexity of surfaces, factor fidelity, relative difficulty).

Because this was a preliminary effort, the experimental apparatus was of only moderate complexity, and the factors manipulated are not directly relevant to issues of simulator design. However, much information obtained in the course of this work has been used to prepare for an experiment now underway on the Naval Training Equipment Center's Visual Technology Research Simulator (VTRS). That work, investigating seven design, subject and difficulty factors, will be published in 1981 as NAVTRAEQUIPCEN 78-C-0060-12.

Stanley C. Collyer

STANLEY C. COLLYER
Scientific Officer

SUMMARY

> In designing the experiment, therefore, the aim of the experimenter should usually be, not to provide for the highest possible degree of precision in the estimate, but rather to secure, <u>with the minimum expenditure of his resources</u>, whatever degree of precision and freedom from bias is <u>sufficient</u> for his purposes.
>
> E. F. Lindquist, <u>Design and Analysis of Experiments in Psychology and Education</u>, 1953

The Visual Technology Research Simulator of the Research Department, Naval Training Equipment Center, Orlando, Florida, was created to support the research objectives of two groups: the engineers, concerned with improving visual system technology, and the psychologists, concerned with evaluating that technology from the standpoint of training effectiveness. On the one hand, the VTRS serves as a test bed for developing and evaluating new engineering concepts aimed at providing increased visual system performance at lower costs. At the same time, it is a research tool for examining how pilot performance and transfer of training are influenced by a wide range of factors of critical interest to simulation engineers and training specialists (Collyer and Chambers, 1978).

Traditional experiments for measuring performance and transfer of training are costly to conduct and limited in the amount of information they generate concerning the large number of factors of interest in the design of training simulators. New approaches that provide more information at less cost have been tried in recent years, but are still costly by practical standards. For an effective research effort, more economical methods particularly suited to transfer of training research must be developed.

The objectives of this study were (a) to establish relationships among training, test, and transfer scores in the context of a complex manual tracking task, (b) to determine the relative complexities of response surfaces for training, test, and transfer, (c) to demonstrate a new research paradigm for the simultaneous investigation of the effects of a large number of simulator design variables on transfer to different vehicle configurations, and (d) to extract from the available data indications that will enable the transfer effectiveness of simulator characteristics to be estimated with minimum costs.

The data for the study were collected at New Mexico State University (NMSU) by personnel of ILLIANA Aviation Sciences. The experimental task involved the lateral tracking of a target following a constantly changing course generated by a microcomputer and presented by appropriate display symbols indicating target position, aircraft position and attitude, and a predicted aircraft position. Dynamic simulator design variables that were systematically manipulated included: (a) vehicle control order, (b) display lag, (c) tracking mode (pursuit vs. compensatory), (d) prediction time, and (e) control gain. A sixth variable was the number of training trials before transition to one of three transfer vehicle configurations, designated Hard, Central, and Easy.

In all, 80 flight-naive young male adults participated in the experiment. Of these, 48 were assigned to three experimental groups of 16 each and were trained under 48 uniquely different simulator configurations. After training, the three groups, respectively, transferred to Hard, Central, and Easy transfer vehicle configurations. The 48 unique combinations of training and transfer conditions were selected so that the effects of all factors, including transfer vehicle configuration, and all two-factor interactions could be isolated. A 49th data point was added at the center of the experimental space on which eight more participants were trained and subsequently transferred to the Central transfer vehicle configuration. These extra data allowed tests of the adequacy of the second-order model and for curvilinearity. Three control groups of eight participants each were tested on the transfer vehicle configurations without prior training.

The data were analyzed by personnel of the Canyon Research Group in California and Florida and were interpreted by the authors. The study demonstrated a unique and economical means of performing holistic transfer of training research and provided useful information regarding relationships among training and transfer performances under an unprecedented number of conditions. The information obtained is being applied to the design of transfer of training experiments to be conducted in the VTRS laboratory.

This experiment was the first to: (a) study the effects of as many as six equipment and training factors in a single transfer of training experiment, (b) examine a broad spectrum of training vehicle configurations--49--in the same experiment, (c) train only a single participant on each of 48 training vehicle configurations, (d) employ multiple transfer configurations--3--in the same experiment, and (e) provide data suitable for deriving multiple regression equations for estimating the effectiveness of configurations not directly studied.

This multifactor, multivariate economical data collection plan demonstrates the effectiveness of the approach, particularly applicable in the preliminary design stage of a simulator engineering program when the relative merits of a large number of simulator configurations must

be compared to establish basic characteristics. Furthermore, the multifactor design reveals each component's contribution to total simulator transfer effectiveness, permitting sources of negative transfer to be identified even though overall transfer is positive.

The following results would appear to have transfer of training applications provided subsequent investigations demonstrate their generality:

1. The transfer surface was less complex than the training surface.

2. The correlation between training and transfer performance for different design configurations was positive but too weak for predictive purposes.

3. Some factors had large effects in training but not in transfer, whereas some with small effects in training had relatively large transfer values.

4. Transfer was facilitated when the values of certain variables resulted in training conditions that were more difficult than the subsequent transfer criterion conditions.

The cost of conducting this six-factor, three-test-configuration study in terms of subject numbers and hours of data collection was less than one-eighth of that for the same information obtained in a series of smaller experiments using conventional methods, and the data were more generalizable, more precise, and less biased.

PREFACE

A number of persons made major contributions to this experiment.

Louis Corl of ILLIANA Aviation Sciences was reponsible for the hardware and software to generate, display, and control the simulated horizontal steering task and for the automatic data acquisition and reduction. Jan Christopher Hull, Paul M. Simon, and Donald G. Fahrenkrog, also of ILLIANA Aviation Sciences, trained and tested the participants.

The 80 participants were volunteers from the local community of Las Cruces, New Mexico, and from the Department of Psychology, New Mexico State University.

Daniel P. Westra of the Canyon Research Group helped analyze and interpret the experimental data, and Brian Nelson of the Canyon Research Group helped prepare special computer software for some analyses.

Linda Carlson-Wenger of ILLIANA Aviation Sciences processed the manuscript.

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF TABLES

LIST OF TABLES

SECTION I

INTRODUCTION

Transfer research to evaluate pilot training simulators has been underway for more than 30 years. Williams and Flexman (1949) reported the first such experiment in 1949. Still there are few simulator and training-program design principles that experts can agree on, and even those widely accepted cannot be stated with sufficient precision to support choices of design parameters with confidence (Roscoe, 1980, Chapters 15-22; NATO-AGARD, 1980).

There are many reasons for our limited understanding of the dependence of training effectiveness on simulator and training-program design. At best, transfer experiments involving complex simulators and real airplanes are logistically difficult and costly. Furthermore, classical experimental designs (Campbell and Stanley, 1963), while suitable for <u>evaluating</u> particular devices or systems, are less suitable for research intended to <u>identify</u> which of a great many variables in training positively affect performance of the same task in the operational situation.

In the real world of pilot training, the number of potentially important simulator and training variables is disturbingly large. They fall into various classes and often interact to a degree that defies comprehensive analysis. To study this large number of variables using conventional approaches has proved unacceptably expensive, if not impossible. To consider only a few at a time produces results that are inevitably inaccurate when applied to operational situations and, because of the situation-specific nature of human behavior, cannot be generalized with confidence far from the values fixed in the experiment.

The problem of conducting transfer of training studies would be greatly simplified if one could realistically assume that transfer of simulator training to actual flight could be predicted from relative performances in the simulator. However, one can easily conceive of situations in which features that make a simulator easier to fly will not necessarily promote higher transfer to the operational situation. Unfortunately there have been few systematic investigations of the relationship between performance during training and transfer of perceptual-motor skills. Until relationships have been established empirically and can be expressed quantitatively, transfer experiments will be required. Finding practical ways of collecting good information inexpensively regarding the effect of training-simulator design variables is a prerequisite to the collection of such data.

SECTION II

APPROACH

Recognizing that a problem exists is the first step in any scientific inquiry. Because it is evident that the problem of obtaining good multifactor training-simulator design information is both economic and logistic, a "holistic" approach to experimentation has been adopted. The term holistic refers to "a philosophic point of view in the conduct of behavioral experiments that emphasizes the importance of accounting for as many potentially critical variables as possible, whether equipment, environment, subject, or temporal, controlled or uncontrolled" (Simon, 1979, p. 77). Human behavior cannot be isolated to the same degree that chemical compounds can be. It is situation specific and must be defined in terms of the total situation.

The key word in the above definition is "critical." Whatever their number, if critical variables are held constant in an experiment, unless the fixed values are close to those found operationally, findings can be grossly inaccurate when applied to the operational situation. Properly implemented, the holistic approach will yield data that are more precise, less biased, and more generalizable from laboratory to field for far less cost than the traditional elemental approach (Simon, 1979, Chapter II). In an ongoing research program, the savings in time and resources can be considerable.

The use of economical multifactor designs and sequential strategies (Simon, 1973; 1977) can significantly reduce the cost of doing holistic experiments. Recently, this approach was successfully implemented in a study of performance during carrier landing as a function of ten equipment/environment factors (Westra, et al., 1981). However, when many factors must be investigated, these economical designs may still not be sufficiently economical if extended training is involved. That is to say, even with a significant reduction in the number of data points sampling the coordinate space, compared with the number called for by traditional factorial designs, the size of the effort may still be impractical to consider.

Such designs may be impractical for transfer of training studies because, for each data point, one or more subjects must be trained for an extended period and subsequently tested during a transfer period. In some cases, still more data must be collected for control groups. Furthermore, the nature of a transfer study requires that a different subject (or subjects) be tested at each data point, a condition that is not always necessary for performance studies in which each subject may be tested on several conditions.

For these reasons, unmodified economical multifactor designs may still not be sufficiently economical for transfer of training research, but what are the alternatives? An investigator may again be tempted to resort to a series of small studies involving a few factors at a time. But this will not solve the problem since, for any given number of factors and for an equal degree of precision, it is less costly to study them all in a single experiment than in any series of smaller experiments that requires expensive replication for precision.

This does not deny the appropriate use of small experiments in a holistic approach, but they must be done in a manner and context that permit the results of each to be included in the ever expanding overall data base. Merely reducing the number of factors to reduce cost also is not a reasonable solution if all are potentially important and their relative importance is unknown.

Because so little is known about the characteristics of a multifactor transfer of training performance space, or the relationships among training, transfer, and intervening variables, the present study was designed to test a particularly economical transfer of training data collection plan and to use the resulting data to investigate the relationships between training and transfer performance.

SECTION III

THE DATA COLLECTION PLAN

It was decided at the start that an artificial data base would not be appropriate to use. Even with a limited task, it was thought that the characteristic relationships between training and transfer would be more faithfully represented by an actual data base than by a structured, computer-generated base that could only reflect the characteristics introduced by the investigator. While the use of a relatively simple task was accepted as a limitation of this exploratory study, it corresponds to some degree with those of primary interest in the real world and the simulated system should involve variables important in the design of operational simulators.

Simulator

Lateral control of a vehicle having simplified dynamic responses similar to an airplane was programed on the MicroGraphic Simulator of the Behavioral Engineering Laboratory, New Mexico State University. This versatile research facility consists, in part, of an ADAC System 1000 microcomputer, a plasma-panel matrix display, and a three-axis manual flight controller, only the lateral dimension of which was used in this simulation. The system also performs automatic performance scoring and immediate data reduction, partial analysis, and visual display, with subsequent printout on a daisy-wheel printer.

Task

The tracking task involved the relative lateral positions of three symbols (Figure 1) on a plasma-panel display located approximately one meter in front of the seated participant. The three symbols appeared, respectively, as an airplane-like figure viewed from the rear, a target box, and a small predictor circle (approximated by an octagon). Whenever the prediction-time variable was set at zero, the predictor circle was superposed concentrically with the vehicle symbol and appeared as if it represented the body (or fuselage) of an airplane.

Using a control stick located on the right arm of the seat, each participant, regardless of the system configuration, was instructed to manipulate the control to keep the vehicle and the target as close to each other as possible (thus the object was to keep the vehicle directly superposed on the target throughout a trial). The simulated vehicle was maneuvered by right and left deflections of the control stick. If the stick were deflected to the left, the vehicle responded in kind, either banking and moving left or just banking left, depending on whether the display mode were pursuit or compensatory.

The target was driven by a forcing function (or course) generated by the summation of four sine waves, a fundamental sinusoid having a period of 41 sec and its 2nd, 5th, and 13th harmonics. Trials were

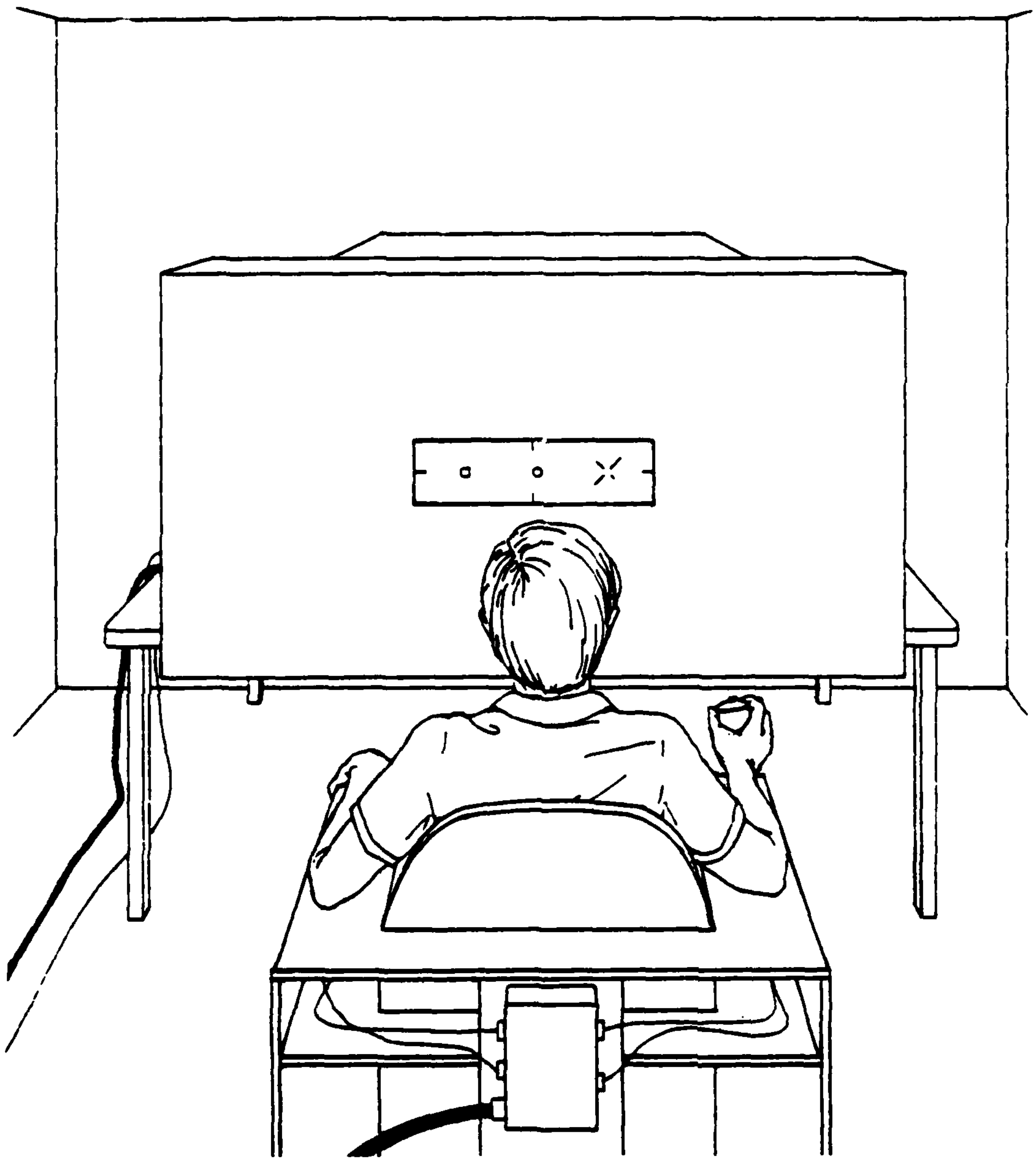


Figure 1. Experimental apparatus.

each 51 seconds in length, of which the first ten seconds were not scored. For any combination of system-configuration variables, the true tracking error between the vehicle and the target was displayed unless one or the other were in saturation at the left or right edge of the screen. Because scoring was based on the displayed error, the operator could see exactly what was being scored.

Experimental Factors

The factors manipulated to form the different simulator configurations were: Control Order (CO), Display Lag (DL), Tracking Mode (TM), Prediction Time (PT), and Control Gain (CG). Number of Training Trials (TT) was a sixth experimental factor. These factors and their levels under training and transfer conditions are listed in Tables 1 and 2, respectively. It was intended that this set of factors would be rich in a variety of transfer characteristics, including high and low performances in training and both high and low positive and negative effects during transfer. This objective was not fully realized. The understanding of the relationships required to make such judgments in advance does not yet exist.

Control Order (CO). The dynamic response of a vehicle to control inputs can range from zero-order to third-order or higher. In zero-order the position of the vehicle (in this case its lateral position) corresponds directly to the deflection of the operator's control stick. In first-order control, stick deflection directly determines velocity, or rate, of lateral movement; in second-order, lateral acceleration; and so it goes. Generally, the higher the order the more difficult the control, but relative transfer will also depend on the closeness of the match between training and transfer control dynamics. For this study, control orders ranged from 100% acceleration (designated the Hard level) to a combination of 75% velocity and 25% acceleration (the Easy level). The midrange (Central) level dynamically combined 38% velocity and 62% acceleration.

Display Lag (DL). Time lags between vehicle response and display indications can have large effects on both performance and transfer. Their negative effect on simulator performance is similar to that of increasing control order, but display lags can be beneficial to transfer if learning to anticipate vehicle responses is a major component of the task, as in formation flying. Typical lags can be exponential in form, as occur when noisy signals are smoothed by filtering, or simple transport delays. Because the latter are common in the updating of digitally generated visual displays in flight simulators, variable transport lag was selected as the second factor. The three levels (-, 0, and +) were 0.30, 0.15, and 0.00 second, respectively.

TABLE 1. TRAINING PHASE: EXPERIMENTAL SYSTEM-CONFIGURATION AND CURRICULUM FACTORS AND THEIR HARD (-), CENTRAL (0), AND EASY (+) VALUES

FACTOR	LEVELS*		
	-	0	+
Control Order (CO) in percent acceleration	100	62	25
Display Lag (DL) in seconds	.30	.15	.00
Tracking Mode (TM) in percent pursuit	0	50	100
Prediction Time (PT) in seconds	.0	.3	.6
Control Gain (CG) in unitless ratios	.12	.18	.24
Number of Training Trials (TT)	10	20	30

*Levels designated - and + are actually coded -1 and +1, respectively, corresponding to what was assumed *a priori* to be the Hard and Easy levels. The 0 is the Central physical position between those limits.

TABLE 2. TRANSFER PHASE: FACTOR VALUES FOR THE HARD, CENTRAL, AND EASY TRANSFER VEHICLE CONFIGURATIONS

CONFIGURATION	FACTORS				
	CO	DL	TM	PT	CG
A. Hard (-)	100	.30	0	0.0	.12
B. Central (0)	62	.15	50	0.0	.18
C. Easy (+)	25	.00	100	0.0	.24

Tracking Mode (TM). Computer-based control and display manipulations can have effects on both performance during training and subsequent transfer of learning. For example, most flight tasks require compensatory control, which appears to be more than twice as inaccurate as pursuit tracking for fairly difficult courses (Roscoe, 1980). In a compensatory tracking mode only the target symbol (Figure 1) moves to indicate tracking error relative to a fixed vehicle index, usually at the center of the display. In a pursuit tracking mode both symbols move to show the absolute positions of the vehicle and the target against fixed display coordinates. By modifying the training system to allow pursuit control, it may be possible that, because correct manual inputs are elicited earlier in training, learning will be faster, and transfer effectiveness higher. For the present study, 100% pursuit displays were taken to be the Easy configuration, and 100% compensatory displays the Hard configuration. The Central configuration was 50% pursuit and 50% compensatory.

Prediction Time (PT). Useful predictors on displays of the type used can be of any order in the Taylor series. For this study, a first-order predictor, showing where the vehicle would be in a few moments with present rate maintained, was represented by a small octagon approximating a circle (Figure 1). In pursuit displays this symbol appeared as a predictor of future vehicle position, while on compensatory displays it appeared as a predictor of the amount by which present error would be changed during the prediction interval. In either case, its position relative to the vehicle symbol represented imminent magnitude of error based on current vehicle velocity. Its three values (-, 0, and +) represented prediction times of 0.0, 0.3, and 0.6 second, respectively.

Control Gain (CG). Control or stick gain is a modifier of the output from the stick to the dynamics of the computer-generated vehicle. Generally speaking, it may be described as the sensitivity of the vehicle to control input. For this experiment, a Central (0) value was selected so that full deflection of the stick in a pure rate-control mode produced vehicle velocity about 20 percent greater than the maximum forcing-function velocity. Thus the vehicle could always overtake the target if the operator applied full stick deflection. If this Central value of gain is regarded as one (or 3/3), then the gain for the Hard configuration was selected as 2/3 and that for the Easy configuration as 4/3. In other words, the gain in the Hard condition was 66% and in the Easy condition 133% of the Central sensitivity. The actual unitless ratio values used for the -, 0, and + conditions were .12, .18, and .24, respectively.

Number of Training Trials (TT). Half of the 48 transfer trainees "flew" 10 training trials (-) and the other half 30 training trials (+) prior to 30 transfer trials in one of three transfer vehicle configurations. The Central group flew 20 trials (0) with the Central training configuration, followed by the usual 30 transfer trials with the Central transfer configuration.

Transfer Configurations

Three different configurations of the simulated transfer criterion vehicle were employed. Their actual compositions are shown in Table 2. One configuration was made up of factor levels to create a Hard task. Another used the factor levels to create an Easy task. A Central configuration was of intermediate difficulty. Centrality was measured on the physical scale of each factor and did not necessarily result in a performance level midway between those attained at the two extremes.

For each transfer configuration there were 16 uniquely different training configurations. Participants who transferred to the Hard configuration after training are referred to as Group A, those transferring to the Central configuration as Group B, and those transferring to the Easy configuration as Group C. A fourth group, D, also transferred to the Central transfer configuration, will be described and its purpose discussed in the section on Data Collection.

The intention here was to increase the generality of our results by creating an experimental space that covered one extreme to the other not only in the training phase but also in the transfer phase. The relationships between training and transfer levels covered extremes of Easy-to-Easy, Easy-to-Hard, Hard-to-Easy, and Hard-to-Hard. Being quantitative factors, any equation developed from the empirical data would be descriptive (and thereby to some extent predictive) of any point within the limits of that experimental space. This was the data base needed to discover training and transfer relationships across a broad multifactor space.

Participants

Eighty right-handed adult male nonpilots with self-reported normal or corrected-to-normal vision participated voluntarily and without pay. The majority of these participants was drawn from the Behavioral Engineering Laboratory's participant roster. These people were volunteers, interested in aviation, recruited from the NMSU campus and local communities. The remainder of the participants came from the Department of Psychology's participant pool and were fulfilling a course requirement by voluntarily participating in this particular experiment.

Data Collection

It was intended to sample performance at a sufficient number of strategically located points in the multifactor space to approximate the complete training and transfer performance surfaces for the variables studied. This goal, however, was subordinated since the data collection effort had a finite limit that did not lend itself to a sequential approach. The data collection plan actually used was sufficient to estimate all main and two-factor interaction effects with enough additional data to estimate how well that limited model fit. Previous experience suggested that this compromise would be adequate.

Performances at 49 coordinate points were sampled in the six-dimensional training space defined earlier. Forty-eight of these points formed a fraction of a 3×2^6 factorial design divided into three orthogonal blocks represented by Groups A, B, and C, who transferred, respectively, to the Hard, Central, and Easy transfer vehicle configurations. From the 16 points within each block, all main effects could be estimated independently from one another and from all two-factor interactions with the interactions aliased (i.e. confounded with each other) in strings. With data from the three blocks combined, all main effects and all two-factor interactions could be isolated from one another, although not from higher-order interactions (Connor and Young, 1961, p. 16).

The 49th point was located at the center of the fractional factorial design and was replicated eight times. Thus, eight participants, designated Group D, were trained at this central position and subsequently transferred to the Central transfer configuration (the same as that for Group B). These centerpoint data provided an estimate of experimental error variance and, when combined with the other data, allowed a test of whether or not a linear second-order model was adequate.*

The coordinates for this data collection plan are given in Appendix A. A different subject was assigned to each of the 48 fractional factorial points and the eight centerpoints, making a total of 56 transfer participants who received the instructions in Appendix B and were then trained and tested in the following sequence:

---

*These centerpoints do not affect the estimates of the effects of the two-level factors. A slightly better design might have been to include two or three Central training conditions within each of groups A, B, and C.

1. Matching trials: Each participant was given three trials in the Central training configuration. The median of these three scores was used to adjust training and transfer scores for individual differences in initial ability.

2. Masking trials: Each participant were then given two trials using his individually assigned training configuration. The purpose of this was to allow any immediate carryover effects from the matching condition to the training condition to dissipate.

3. Training trials: Each participant was given 10, 20, or 30 trials on his specific training vehicle configuration, the number depending on his coordinate position in the six-dimensional training space.

Total instructional and "flying" time for these first three steps of the data collection ranged from approximately 25 to 45 minutes depending on the number of training trials involved.

4. Transfer trials: Following a 30-minute rest period away from the experimental room, each participant was brought back and tested for 30 trials on one of the three transfer vehicle configurations.

Three additional groups of eight participants each also received the instructions in Appendix B and were tested to provide "control" performances on the three transfer vehicle configurations. Each group was trained on a different transfer configuration without prior training on any other configuration except for the matching trials. As with the transfer groups, each control participant was given the three matching trials at the center of the training space and the two masking trials with his particular training vehicle configuration prior to the 30 scored training trials.

SECTION IV

MAXIMIZING AND EVALUATING DATA QUALITY

As experimental designs become more complex and capable of generating large quantities of information from relatively few data points, a necessary first step in the analysis is to examine the quality of the data. This is imperative to ensure its valid interpretation. The data in this study were examined (a) to decide on proper analyses, and (b) to detect characteristics of the data that might distort interpretation. Rather than accept the numbers uncritically, the following questions were asked:

1. In what form are the data most suitable for analysis?
2. Were the assumptions of control and balance in the experimental design actually met?
3. Are the data distorted by any experimental procedure?
4. Do the data behave as expected?
5. Are there internal inconsistencies in the data?
6. Are observed outliers meaningful or consequences of poor data collection techniques?

Analyses were performed to answer these and similar questions. The more important conclusions are listed below:

1. To provide a more normal distribution of the performance data, logarithmic transformations were made of the original RMS error scores. All further discussion of performance scores will be in terms of the log RMS values.
2. In combining the results from several trials into a single measure, the median value (rather than the mean) was used.
3. The seven groups (A, B, C, D, and three controls) were well-matched in terms of the means and variances of their matching scores.
4. Although the data were collected by three experimenters over a period of several weeks and at different times of day and evening, there was no evidence that these sources of variance biased the data to any practical extent.

5. The amount of learning during training was not large, but the overall curves showed expected patterns. Individual curves, however, were characterized by large variations, both within and between configurations.

6. The assumptions made in selecting the Hard and Easy levels for the factors and the transfer configurations were validated. Only Prediction Time did not conform to expectations.

Supporting data regarding these conclusions are given in Appendix C.

SECTION V

TRAINING AND TRANSFER: SURFACES AND EFFECTS

The data were analyzed to obtain information regarding the training and transfer surfaces as well as the individual effects. To describe the training and transfer performance surfaces, data from the 56 participants in the four experimental groups were used. When transfer effects are considered, these values were obtained by adjusting each individual's transfer performance score by the mean performance level of the control group tested with the same transfer vehicle configuration.

Individual Initial Abilities

Prior to training, each of the 80 participants was tested for three trials on the same (Central) system configuration. The median of these three trials was used to represent each participant's initial ability or skill level and is referred to as his matching score. The mean of the matching scores for the 80 participants was 1689.44, with a standard deviation of 82.15 and a standard error of the mean of 9.18. The smallest value was 1520, approximately two standard deviations below the mean; the largest value was 1912, approximately two and three-quarters standard deviations above the mean.

Differences among the individual's initial abilities, as measured by the matching scores, were partialed out of all training and transfer scores.* All subsequent references to training performances, transfer performances, and transfer effects are based on the adjusted values from which that portion of a score attributable to initial individual ability has been partialled out.

Interpreting Signs

Throughout this paper, plus and minus signs are used to represent the coded levels of the factors on the input side of the experiment and the signs of the factor effects (twice the coefficients of the regression equations) and transfer effects scores on the output side. To facilitate the interpretation of these signs, Table 3 is provided here, to be used when needed to understand the results throughout the remainder of this paper.

---

*Caution: With designs in which conditions and subjects are confounded, it is important that scores of the matching task clearly reflect individual differences in ability. Prevalidation of the matching task is paramount.

TABLE 3. LEGEND FOR INTERPRETING SIGNS

INPUTS

Coded levels of equipment/training factors

- -1 Preexperiment, assumed Hard (High log RMS error)
- +1 Preexperiment, assumed Easy (Low log RMS error)

Coded levels of similarity variables:

Direction (Difficulty)

- -1 Hard to Easy
- 0 No change
- +1 Easy to Hard

Distance (Fidelity)

- -1 Same position (0)
- 0 Moved 1 position
- +1 Moved 2 positions

OUTPUTS

Factor Coefficients (Coeff. x 2 = Mean difference = Effect) for training and transfer scores

- \- Hard Level had larger RMS error than Easy Level
- \+ Easy Level had larger RMS error than Hard Level

Transfer effects scores (observed minus control scores)

- \- Positive Transfer (Control error higher than observed)
- \+ Negative Transfer (Control error lower than observed)

Similarity Coefficients (Coeff. x 2 = Mean difference = Effect)

Direction (Difficulty)

- \- Hard-to-Easy error greater than Easy-to-Hard, or Hard-to-Easy had poorer transfer
- \+ Easy-to-Hard error greater than Hard-to-Easy, or Hard-to-Easy had better transfer

Distance (Fidelity)

- \- Shorter distance had larger error than Longer distance, or Shorter had poorer transfer
- \+ Longer distance had larger error than Shorter distance, or Longer had poorer transfer

Training Data

The median scores for the last five training trials, regardless of the total number of trials, were analyzed for the training surface characteristics and for the effects of the individual factors and their interactions. These analyses were applied to the combined set of 56 scores for Experimental Groups A, B, C, and D and separately to each of the three sets of 16 scores for Groups A, B, and C who subsequently transferred to the Hard, Central, and Easy vehicle configurations. Summaries of the analyses for surface characteristics of the training performance data are shown in Tables 4, 6, 8, and 10 and for the factor effects in Table 12.

Transfer Data

Median performances on the first five trials with the respective transfer vehicle configurations were used for all participants, both those trained and the control groups. Transfer performance scores for the 56 trained operators were analyzed in the same manner as the training data. To estimate the transfer effect scores, group mean performance scores for each of the three control groups were subtracted from the corresponding individual transfer performance scores of the experimental groups. Thus, for individual measures, positive transfer savings are indicated by negative difference scores (large control scores subtracted from smaller transfer scores), and conversely negative transfer effects are indicated by positive difference scores.

Summaries of the analyses of surface characteristics of the transfer performance data are shown in Tables 5, 7, 9, and 11, those for the factor effects in Table 13. The analysis of the transfer effects is not shown but will be covered in the discussion of Table 13. See Appendix D regarding the percent transfer analysis.

Training and Transfer Surface Characteristics

When Tables 4 and 5 are examined and compared, the following generalizations can be drawn:

1. Both training and transfer surfaces can be approximated predominantly by a combination of first-order effects.

2. The training surface is somewhat more complex than the transfer surface, being influenced more by two-factor interactions and showing a somewhat greater lack of fit.

3. Evidence of curvature is slight in both surfaces.

4. Experimental factors have a major influence on performance during training but trivial effects on transfer in the presence of the large performance differences imposed by the three transfer vehicle configurations.

TABLE 4. ANALYSIS OF TRAINING PERFORMANCE SURFACE: GROUPS A, B, C, AND D

Source of Variance	df	Mean Squares	F	< p	Proportion of Variance	
Regression:	35	71531	79.0	.000*	.94	
First-order terms	8	261126	288.5	.000	.79	
2 Levels	6	332868	367.8	.000		.75
3 Levels	2	45902	50.7	.000		.03
Second-order terms	27	15354	16.9	.000	.16	
2 FI (2 x 2 levels)	15	17201	19.0	.000		.10
2 FI (2 x 3 levels)	12	13406	14.4	.001		.06
Residual:	20	7461			.06	
Lack of Fit	12	4742	5.2	.025	.02	
Curvature	1	85981	95.0	.000	.03	
Error	7	905			.00	

TABLE 5. ANALYSIS OF TRANSFER PERFORMANCE SURFACE: GROUPS A, B, C, AND D

Source of Variance	df	Mean Squares	F	< p	Proportion of Variance	
Regression:	35	190125	43.7	.000*	.97	
First-order terms	8	744636	163.3	.000	.87	
2 Levels	6	26181	5.7	.025		.02
3 Levels	2	2900002	636.1	.000		.84
Second-order terms	27	25826	5.7	.025	.10	
2 FI (2 x 2 levels)	15	24794	5.4	.025		.05
2 FI (2 x 3 levels)	12	27116	5.9	.025		.05
Residual:	20	11267			.03	
Lack of Fit	12	5558	1.3	.10	.00	
Curvature	1	128170	29.4	.010	.03	
Error	7	4354			.00	

*p = .000 is < .0005.

TABLE 6. ANALYSIS OF TRAINING PERFORMANCE SURFACE: GROUP A

Source of Variance	*df*	Mean Squares	*F*	<*p*	Proportion of Variance
First-order terms	6	134475	8.6	.005	.85
Residual	9	156635			.15
Second-order terms (aliased)	7	18085	2.6	.10	.13
Remainder	2	7061			.01

TABLE 7. ANALYSIS OF TRANSFER PERFORMANCE SURFACE: HARD VEHICLE CONFIGURATION (GROUP A)

Source of Variance	*df*	Mean Squares	*F*	<*p*	Proportion of Variance
First-order terms	6	10304	0.7	.10	.31
Residual	9	15633			.69
Second-order terms (aliased)	7	17500	1.9	.10	.60
Remainder	2	9101			.09

TABLE 8. ANALYSIS OF TRAINING PERFORMANCE SURFACE: GROUP B

Source of Variance	*df*	Mean Squares	*F*	$<p$	Proportion of Variance
First-order terms	6	106897	9.6	.005	.87
Residual	9	11085			.13
Second-order terms (aliased)	7	11162	1.0	.10	.11
Remainder	2	10816			.03

TABLE 9. ANALYSIS OF TRANSFER PERFORMANCE SURFACE: CENTRAL VEHICLE CONFIGURATION (GROUP B)

Source of Variance	*df*	Mean Squares	*F*	$<p$	Proportion of Variance
First-order terms	6	55681	3.2	.06	.68
Residual	9	4453			.32
Second-order terms (aliased)	7	4913	1.7	.10	.27
Remainder	2	2845			.04

TABLE 10. ANALYSIS OF TRAINING PERFORMANCE SURFACE: GROUP C

Source of Variance	*df*	Mean Squares	*F*	< *p*	Proportion of Variance
First-order terms	6	117485	11.1	.001	.88
Residual	9	10616			.12
Second-order terms (aliased)	7	10716	1.0	.10	.09
Remainder	2	10268			.03

TABLE 11. ANALYSIS OF TRANSFER PERFORMANCE SURFACE: EASY VEHICLE CONFIGURATION (GROUP C)

Source of Variance	*df*	Mean Squares	*F*	< *p*	Proportion of Variance
First-order terms	6	55681	1.3	.10	.47
Residual	9	42623			.53
Second-order terms (aliased)	7	42168	1.0	.10	.41
Remainder	2	44216			.12

TABLE 12. ANALYSIS OF TRAINING PERFORMANCE EFFECTS: GROUPS A, B, C, AND D

Source of Effect*	Regression Coefficient**	Standard Error	Standardized Coefficient	p
INTERCEPT	1710.9			
Control Order	-123.5	12.5	-.53	.000
Display Lag	-157.9	12.5	-.67	.000
Tracking Mode	-36.5	12.5	-.16	.008
Prediction Time	2.7	12.5	.01	.831
Control Gain	-2.0	12.5	-.01	.872
Training Trials	-8.8	12.5	-.04	.488
TVC, Linear (XL)	-47.2	15.3	-.16	.006
TVC, Quadratic (XQ)	16.6	10.0	.09	.113
CO x DL	-38.7	15.3	-.17	.020
CO x TM	-9.1	13.2	-.04	.500
CO x PT	13.1	13.2	.06	.335
CO x CG	-4.0	13.2	-.02	.764
CO x TT	6.6	13.2	.03	.622
CO x XL	3.5	15.3	.01	.820
CO x XQ	11.5	8.8	.07	.206
DL x TM	-29.9	13.2	-.13	.035
DL x PT	39.9	13.2	.17	.007
DL x CG	-12.6	13.2	-.05	.353
DL x TT	6.2	13.2	.03	.644
DL x XL	-3.6	15.3	-.01	.814
DL x XQ	-19.6	8.8	-.12	.038
TM x PT	19.6	13.2	.08	.155
TM x CG	25.2	15.3	.11	.115
TM x TT	-10.2	13.2	-.04	.452
TM x XL	-29.9	15.3	-.10	.065
TM x XQ	5.9	8.8	.04	.509
PT x CG	-19.9	13.2	-.09	.148
PT x TT	22.0	15.3	.09	.165
PT x XL	6.8	15.3	.02	.662
PT x XQ	-14.3	8.8	-.09	.119
CG x TT	-9.6	13.2	-.04	.477
CG x XL	-2.2	15.3	-.01	.887
CG x XQ	10.4	8.8	.06	.250
TT x XL	-20.3	15.3	-.07	.198
TT x XQ	-17.4	8.8	-.11	.063

*Sources that were included in an equation from a stepwise regression analysis with F = 4.00 to enter and exit are underlined.

**Two times the Regression Coefficient equals the mean difference between + and - levels. For interactions, the mean difference is between (++ and --) and (+- and -+).

TABLE 13: ANALYSIS OF TRANSFER PERFORMANCE EFFECTS: GROUPS A, B, C, AND D

Source of Effect*	Regression Coefficient**	Standard Error	Standardized Coefficient	p
INTERCEPT	1688.2 (-11.1)			
Control Order	-14.5	15.3	-.04	.355
Display Lag	-32.4	15.3	-.09	.047
Tracking Mode	31.7	15.3	.08	.052
Prediction Time	6.5	15.3	.02	.678
Control Gain	31.1	15.3	.08	.056
Training Trials	0.5	15.3	.00	.975
TVC, Linear (XL)	-408.8 (-70.3)	18.8	-.88	.000
TVC, Quadratic (XQ)	78.0 (+31.7)	12.3	.26	.000
CO x DL	4.2	18.8	.01	.825
CO x TM	-20.9	16.3	-.06	.213
CO x PT	22.4	16.3	.06	.184
CO x CG	13.2	16.3	.04	.427
CO x TT	11.7	16.3	.03	.480
CO x XL	-27.6	18.8	-.06	.157
CO x XQ	10.4	10.8	.04	.350
DL x TM	-18.2	16.3	-.05	.277
DL x PT	27.6	16.3	.07	.105
DL x CG	2.2	16.3	.01	.895
DL x TT	35.7	16.3	.09	.040
DL x XL	-10.3	18.8	-.02	.588
DL x XQ	7.0	10.8	.03	.524
TM x PT	-39.1	16.3	-.10	.026
TM x CG	-5.3	18.8	-.01	.782
TM x TT	18.8	16.3	.05	.260
TM x XL	-58.2	18.8	-.13	.006
TM x XQ	25.5	10.8	.09	.029
PT x CG	27.4	16.3	.07	.107
PT x TT	-20.2	18.8	-.05	.296
PT x XL	22.4	18.8	.05	.246
PT x XQ	-14.9	10.8	-.06	.185
CG x TT	-46.7	16.3	-.12	.009
CG x XL	-20.0	18.8	-.04	.300
CG x XQ	24.7	10.8	.09	.034
TT x XL	10.0	18.8	.02	.599
TT x XQ	-.7	10.8	-.00	.952

*Sources that were included in an equation from a stepwise regression analysis with F = 4.00 to enter and exit are underlined.

**Two times the Regression Coefficient equals the mean difference between + and - levels. For interactions, the mean difference is between (++ and --) and (+- and -+). Coefficients in parentheses are those of the transfer effects that were different from transfer performance effects (see page 31).

When the data for each group associated with a transfer vehicle configuration are examined separately (Tables 6, 7, 8, 9, 10, 11) the following generalizations can be made:

1. The combined first-order effects of the experimental factors account for most of the variation in performance during training for each group. (Note: Transfer vehicle configurations were not a source of variance during training.)

2. No clear pattern can be seen in the results of the transfer data across the three configurations. Effects are marginal or trivial. Compared to the training surface, the transfer surface is relatively flat.

Interpretation of Training and Transfer Effects

The contributions of the individual sources of variance, namely 35 main and two-factor interaction effects, during training and transfer are shown in Column 2 of Tables 12 and 13, respectively.* Each coefficient shows the change in performance per unit change in the corresponding source of variance. Since performance is measured in log RMS error, a coefficient with a negative sign indicates that performance was poorer on the (-1) level, designated Hard, of the factor associated with the coefficient. The mean difference between the easy and hard levels of any factor can be obtained by multiplying its coefficient by two. For two-factor interactions, the difference is between the means of those values in which both factors are of the same sign (++ and --) and those in which they have different signs (+- and -+).

Standardized regression coefficients (SRCs) in Column 4 are calculated from scores normalized in units of their own standard deviations about the mean. SRCs squared approximate the proportion of total variance accounted for by each source after other sources have been partialled out. The probability that an individual coefficient of the size shown in Column 2 would occur by chance is provided (for a two-tailed _t_-test) in Column 5. A _p_ value of ".000" is actually any value smaller than .001. Each source accounts for one degree of freedom; for these calculations, there were 20 degrees of freedom in the error term.

---

*The three-level factor, Transfer Vehicle Configuration (TVC), was divided into two orthogonal components, namely the linear and quadratic trend effects (XL and XQ, respectively) with one degree of freedom each. XL provides a comparison of the two end levels, and XQ compares the mean of the two end levels with the middle level. These interact with the other factors in the conventional manner.

The underlined Sources of Effects in Tables 12 and 13 indicate those terms that emerge from a stepwise regression analysis in which the criterion for entry and exit is an *F*-value of 4.00. The coefficients for these terms in the stepwise analysis will be the same as those in the full regression analysis except when one source overlaps (interacts with) others. In the full analysis, the reported coefficients are corrected for overlap. In the stepwise analysis, no correction is made unless both interactions in an overlapping pair are brought into the equation. Thus, the differences that may exist for some terms are slight and not relevant to this discussion.

If we examine Tables 12 and 13, certain results stand out clearly:

1. Factors CO (Control Order), and DL (Display Lag), show relatively large effects during training but small or marginal effects during transfer.

2. In transfer, the three Transfer Vehicle Configurations (Factors XL and XQ) overwhelm all other sources of variance (this was as intended, of course), with the next largest sources being interactions (TM x XL, CG x TT, TM x PT) rather than main effects, and these contributed only 10 or 12 percent as much variance as the linear Transfer Vehicle Configurations effect (XL).

3. It would appear that fewer than ten of the 35 isolated sources of variance in either group had a critical effect on performance, quite in line with the Principle of Maldistribution (Simon, 1973, 1977), a fundamental assumption in economical multifactor research.

Selecting which effects are truly critical is difficult in this study because there are no real-world criteria to use. Ordinarily one would look at the mean differences first to see which effects approach operationally critical values. Then one would examine the proportions of variance accounted for and tests of statistical significance. These three criteria are not necessarily consistent and require careful interpretation. Using the *p*-values in Column 5 as the basis for selection requires consideration of the number of effects being examined, in this case 35. Since proper selection of critical factors depends in part on real system requirements, there are no good rules for deciding in this artificial situation.

In Tables 14 and 15, the results from stepwise regression analyses of Groups A, B, and C individually associated with the three vehicle configurations during transfer are shown for both the training and transfer data. In all cases, an *F* of 4.00 was used as the criterion for entry to or exit from the equation. While these analyses are based on only 16 observations each, they allow an examination of the results within transfer configurations rather than between transfer configurations and thereby eliminate comparison of interactions of other factors with XL and XQ.

TABLE 14. STEPWISE REGRESSION ANALYSIS OF TRAINING PERFORMANCE WITH F = 4.00 TO ENTER AND EXIT

Group A

Equation: $1774.67 - \overset{(.27)*}{115.53}\ CO - \overset{(.60)}{173.86}\ DL + \overset{(.03)}{40.83}\ (CO \times TT + DL \times PT)$

Adjusted R^2 = .880 F = 37.80 (3/12), $p < .001$

S.E. of Coefficient = 19.97 S.E. of Estimate = 78.90

Group B

Equation: $1725.73 - \overset{(.46)}{146.61}\ CO - \overset{(.30)}{118.64}\ DL$

Adjusted R^2 = .732 F = 21.50 (2/13), $p < .001$

S.E. of Coefficient = 28.76 S.E. of Estimate = 115.04

Group C

Equation: $1680.27 - \overset{(.20)}{108.50}\ CO - \overset{(.55)}{181.14}\ DL$

Adjusted R^2 = .715 F = 19.80 (2/13), $p < .001$

S.E. of Coefficient = 33.56 S.E. of Estimate = 134.23

---

*Numbers in parentheses above factors indicate proportions of total variance accounted for.

TABLE 15. STEPWISE REGRESSION ANALYSIS OF TRANSFER PERFORMANCE WITH $\underline{F}$ = 4.00 TO ENTER AND EXIT

Hard Vehicle Configuration (Group A)

$$\text{Equation: } 2174.91 + \overset{(.30)^{*}}{115.35\ \text{TM}}$$

Adjusted $\underline{R}^2$ = .246 — $\underline{F}$ = 5.90 (1/14), $\underline{p}$ < .05

S.E. of Coefficient = 47.47 — S.E. of Estimate = 189.89

Central Vehicle Configuration (Group B)

$$\text{Equation: } 1591.14 - \overset{(.27)}{46.47\ \text{DL}}$$

Adjusted $\underline{R}^2$ = .22 — $\underline{F}$ = 5.24 (1/14), $\underline{p}$ < .05

S.E. of Coefficient = 20.30 — S.E. of Estimate = 81.21

Easy Vehicle Configuration (Group C)

$$\text{Equation: } 1357.34 + \overset{(.31)}{62.80\ (\text{CO} \times \text{DL} + \text{DL} \times \text{TT})}$$

Adjusted $\underline{R}^2$ = .26 — $\underline{F}$ = 6.33 (1/14), $\underline{p}$ < .025

S.E. of Coefficient = 24.95 — S.E. of Estimate = 99.81

---

*Numbers in parentheses above factors indicate proportions of total variance accounted for.

Important sources of variance. Control Order and Display Lag (CO and DL) have strong effects on the training performance of each group. These two factors accounted for more than 70% of the variance in each analysis. With the transfer data, the picture is different. No more than a single factor or interaction string meets the criterion for entry into the transfer equation for any group, and entries are not consistent from group to group. Terms admitted to each of these equations account for only about 30% of the total variance. There is not sufficient information to decide why a particular factor affected a particular transfer configuration, nor to determine what caused the unexplained variability.

Transfer effects. The analyses of transfer effects scores yield the same coefficients as those for the transfer performance scores, shown in Table 13, with three exceptions: the intercept, XL, and XQ. These three values are listed in parentheses on the proper lines in Column 2. The reason for this is that when the average performance of the appropriate control group is removed from each transfer performance score, the differences among transfer vehicle configuration levels (XL and XQ) are diminished accordingly. However, the resulting "transfer effect" scores will be affected in exactly the same way by the other experimental factors as the transfer performance scores are, and by the same amount. See Appendix D regarding percent transfer scores.

Relations between training and transfer. To facilitate examination of factor effects across training and transfer boundaries, the 48 performance measures of groups A, B, and C from both phases were combined. The 96 scores were subjected to a stepwise regression analysis (F = 4.00 to enter and exit) with all the main effects and two-factor interactions also interacting with a new factor, termed experiment Phase (P), which has training (-1) and transfer (+1) as its two levels. This enabled the effects of 71 sources of variance to be isolated (with Transfer Vehicle Configuration divided into linear and quadratic sources). There were nine main effects, 35 two-factor interactions, and 27 three-factor interactions to be considered.

Based on the criterion indicated earlier, 17 of the 71 possible terms were used in the equation. In combination, these accounted for 91% of the total variance of the 96 scores. These terms are listed in order of their strength of effect on performance in Table 16 where component XL and XQ terms have been combined into the single Transfer Vehicle Configuration variable, X (with two degrees of freedom). The more interesting of these have been plotted in Figure 2 and will be discussed below. Those of greatest interest are the two- and three-factor interactions that involve both training and transfer phases.

TABLE 16. RESULTS FROM STEPWISE REGRESSION ANALYSIS OF COMBINED TRAINING AND TRANSFER PERFORMANCE DATA WITH $\underline{F}$ = 4.00 TO ENTER AND EXIT (N = 96)

Source of Effect		Regression Coefficient	Proportion of Total Variance
T V Configuration (X)	Linear (L)	-227	.376
	Quadratic (Q)	+29	
Phase* x X	L	-181	.242
	Q	+29	
Display Lag		-95	.094
Control Order		-69	.050
Display Lag x Phase		+63	.041
Control Order x Phase		+55	.031
Tracking Mode x X	L	-44	.032
	Q	+16	
Display Lag x Prediction Time		+37	.014
Tracking Mode x Phase		+34	.012
Display Lag x Training Trials		+27	.007
Control Gain x Training Time		-25	.006
Tracking Mode x Prediction Time x Phase		-23	.006
Display Lag x Tracking Mode		-23	.005
Control Gain x XQ		+18	.005

*Phase refers to Training Trials vs. Transfer Trials.

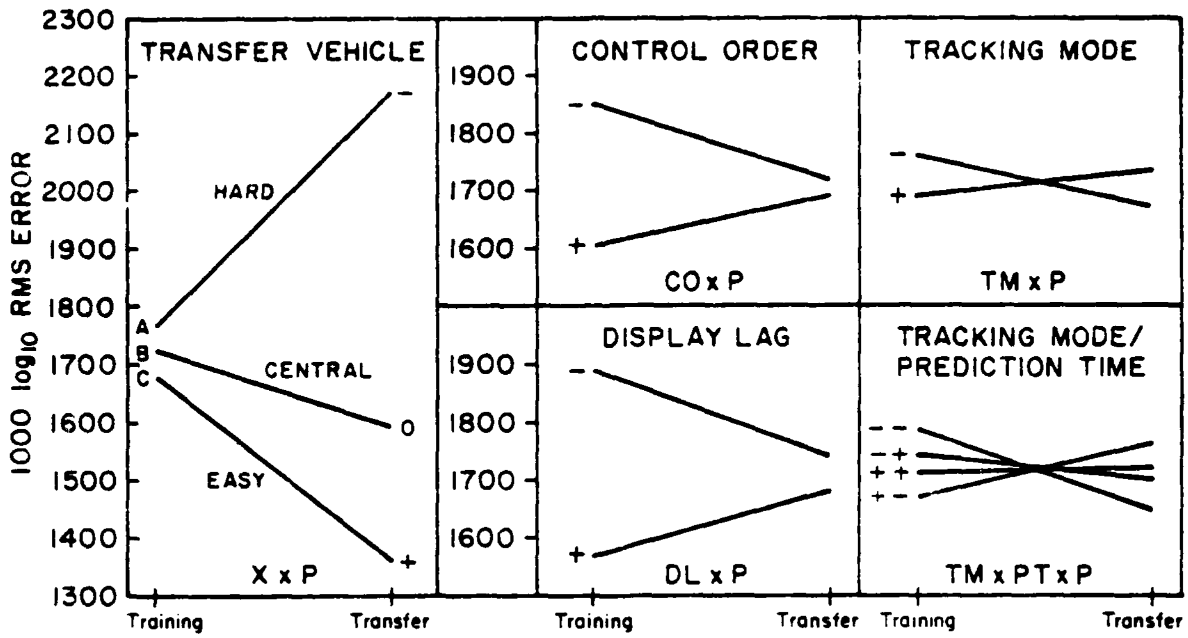


Figure 2. Critical interactions between equipment/training factors and phases.

The three largest sources of variance in this analysis are still the main effects of Factors X (Transfer Vehicle Configuration), CO (Control Order), and DL (Display Lag), but this observation is simplistic and misleading. This is so because all three of these sources also show strong interactions with the phase of training. The plots of X x P, CO x P, and DL x P are shown in Figure 2. These help us understand what the numbers mean. The plot of the XP interaction shows exactly what the investigators had purposefully designed into the experiment.

In transfer, the three widely different transfer vehicle configurations had large differential effects; in training, where 48 unique configurations were employed, with levels of individual factors balanced both within and across groups, differences among the three groups were small, as they should be. This combination creates a numerical X x P interaction. The large overall X x P effect is due to the fact that the averaging of the small group differences during training and the large group differences during transfer still resulted in a large overall difference, but it is meaningless in the light of the interaction and our knowledge of the experimental design.

Our interest in the large CO and DL effects is overridden by the existence of non-trivial CO x P and DL x P interactions. Both CO and DL show large effects during training and relatively small effects during transfer. These are the interaction effects.

Actually some of the smaller effects in this combined analysis are of greater interest; the most important of these are the ones involving an interaction with P. For a factor not to interact with P means that its effect is essentially the same during the training and transfer phases. When a factor does interact with P, it means that a marked change took place between training and transfer. It is necessary to determine whether this change is just a change in the magnitude of an effect (as in the cases of X, CO, and DL) or whether an intrinsic interaction exists in which the effect of a factor reverses itself between training and transfer. Only TM (Tracking Mode) and PT (Prediction Time) appear with P (Experiment Phase) in an intrinsic interaction.

In Figure 2 the TM x P and the TM x PT x P interactions are plotted. In both cases, classic intrinsic interaction can be seen. The rank orders of performance are inverted as a function of whether the measures were for the training or transfer phase. The numbers tell us that the effects, though small, are probably reliable. In a simulator-to-airplane experiment, this result could be important if valid, and we would have independent criteria with which to evaluate the degree of importance. Even more pertinent, such a finding would encourage support for the investigator to extend his examination of the phenomenon.

In Table 17, all sources of variance are listed that show intrinsic interaction effects, i.e., a reversal in the signs of their coefficients from training to transfer. With the exception of the two with asterisks, none of these is statistically reliable at the level used for the stepwise regression equation (namely, $\underline{F} = 4.00$ to enter and exit). These results are shown here merely as a matter of interest, since whether they would be important or not depends on the magnitude of the effect and not on the results of a statistical test or the proportion of variance accounted for.

TABLE 17. REVERSALS IN PERFORMANCE LEVELS AS A FUNCTION OF THE EXPERIMENTAL PHASE

Factors and factor interactions for which the training coefficient is larger than the transfer coefficient:

	Training Coefficient was Negative; Transfer was Positive			Training Coefficient was Positive; Transfer was Negative	
	Training Trials			Tracking Mode x Control Gain	
..	Control Order x Display Lag		.	Prediction Time x Training Trials	
.	Training Trials x TVC, Linear				
..	Display Lag x TVC, Quadratic				
..	Tracking Mode*	..			

Factors and factor interactions for which the training coefficient is smaller than the transfer coefficient:

	Training Coefficient was Negative; Transfer was Positive			Training Coefficient was Positive; Transfer was Negative	
	Control Gain	..		Control Order x TVC, Linear	.
	Control Order x Control Gain		.	Tracking Mode x Prediction Time*	..
	Tracking Mode x Training Trials	.			
	Prediction Time x Control Gain	.			

*Effect x Phase interaction was statistically significant.
.Number of dots indicates how much greater in size (in whole units) each coefficient was than its standard error. Those to left of term are for the training coefficients and to the right, transfer coefficients.

SECTION VI

ECONOMICAL APPROACHES TO TRANSFER OF TRAINING RESEARCH

One purpose of this study was to discover relationships between training and transfer performance that might be generalizable to transfer of training experiments of the type to be performed at the VTRS laboratory. Another purpose was to provide an opportunity to manipulate the limited data base to discover more economical approaches to transfer of training experiments. Of these two goals, the latter was achieved more successfully than the former. A far simpler task was employed in the experiment than would be expected in any simulator likely to be of interest. This probably resulted in a less rich data base than might have been desired. As a result, generalizing results regarding training-transfer relationships must be done cautiously and tentatively.

On the other hand, the test of a new economical approach to transfer of training research is exciting and can immediately promise better and less expensive information when applied to more complex simulation problems. After 35 years of simulation experiments using conventional transfer designs, this was the first to examine more than a few equipment/training factors and multiple transfer vehicle configurations in the same experiment. The information obtained regarding both the effectiveness and problems of this experimental plan is being employed in the planning of a new transfer of training experiment at the Naval Training Equipment Center.

Comparing Present Approach with Conventional Approaches

This study illustrates quite clearly the tremendous economy that can be achieved with multifactor designs that ordinarily obtain equivalent or better information than conventional designs. The design provided the data needed to describe the relationships among six equipment/training factors and three transfer vehicle configurations using only 48 data collection points. Eight additional data points were added to estimate the lack of fit, curvilinearity, and error. Twenty-four more measures were made, making a total of 80 in all, to obtain estimates of the average performance of three control groups with eight subjects each.

In the design, 49 individual training configurations were examined and an equation was obtained that would approximate performance on any combination of training and transfer configurations within the experimental space. Although only one subject per training configuration was used in the basic design, the mean of each level of every two-level factor in the design was based on 24 measurements. The means of each of the three transfer vehicle configurations were based on 16 measurements. How does this compare with more conventional approaches?

One factor at a time. Suppose the investigator had looked at the effect of one two-level training factor on one transfer configuration at a time. The design for each factor might look like this:

	Matching	Training*	Transfer
Experimental Group 1	X_m	(1)	X_t
Experimental Group 2	X_m	(a)	X_t
Control Group	X_m		X_t

Direct comparison of precision between the one-factor study and the multifactor experiment is not easy to do without knowledge of the error variance of the two studies. Still some limiting calculations can show how expensive the one-factor-at-a-time approach can be, in both information and resources, when compared with the multifactor approach. For example, instead of requiring that the mean of each level of a factor be based on 24 observations as in our multifactor experiment, let us require only 12 observations (subjects) per experimental group in the one-factor study. Thus each factor per transfer configuration would require 24 measurements for the two experimental groups.

With six factors and three transfer configurations, we would need 24 x 18 = 432 measurements, plus 24 more for eight subjects per the three control groups, or a total of 456. For this conservative number, almost six times larger than the current study, we obtain less precise estimates of means (and mean differences) and no estimates of any interaction effects. Also an often overlooked point is that since each single-factor study holds the other factors at a fixed level, the generalization of the findings to a wide range of conditions is severely limited.

*Letters are the conventional coding of each experimental condition, where the presence of a letter indicates the high level of that factor and the absence of a letter indicates the low level of the factor. The (1) indicates that all factors are at their lower level.

Two factors at a time. Suppose the investigator decides to test two factors at a time to obtain both main and two-factor interaction estimates for each of the three transfer configurations. Then the design for a single pair of factors and one transfer configuration might look like this:

	Matching	Training	Transfer
Experimental Group 1	X_m	(1)	X_t
Experimental Group 2	X_m	(a)	X_t
Experimental Group 3	X_m	(b)	X_t
Experimental Group 4	X_m	(ab)	X_t
Control Group	X_m		X_t

To compare the costs of this approach with those of the present study, let us assign six subjects to each of the experimental groups. This again probably produces less precise estimates of main and interaction effects than were obtained here. With six factors, there are 15 different pairs of two-factor interactions. To collect the data required to estimate all two-factor interactions for three transfer configurations, we would need six subjects times four experimental groups times three transfer configurations, or 72 for each pair of factors, times 15 pairs for a total of 1080 data points. To this we add 24 more subects with eight used in each of the three control groups. This makes a grand total of 1104 points.

With the two-factor approach, we obtain estimates of all main and two-factor interaction effects, as with the present study, but this approach requires nearly 14 times the effort that the present study did. Furthermore, the two-factor approach is probably less precise in its estimates and surely more biased because the factors not included in two-factor studies are held constant (see Simon, 1979, Section II).

The present design. It can be seen that as long as the same information regarding main and two-factor interaction effects is required, the design used will ordinarily be cheaper and more precise and its results less biased and more generalizable than comparable few-factors-at-a-time approaches.

The greatest advantage of the present design, however, is that it ties the information together in a single equation of specified precision; if the investigator needs more precision, he must collect more data. Rather than looking at isolated segments of the total space, this design provides a description over the entire multifactor space within which interpolations can be made with some confidence. This means that the data can be generalized beyond the specific experimental

conditions included in the design, and it is easier to "hang" new data on this original frame when new factors are studied. This permits a modular data base to be constructed.

Indications Regarding Transfer of Training Experiments

As intended, the present study provides some empirical data that can be useful in planning future transfer of training experiments. However, since this is the first study of its kind, involving a relatively simple task and with little learning and limited transfer, one must consider the following observations as "indications," a concept of data analysis that Mosteller and Tukey (1977) bring to a high level of respectability.*

In this experiment, however, several results were observed that are quite important in the design of transfer of training studies provided they are generalizable:

1. The transfer performance surface appears less complex than the training performance surface (see Tables 4 and 5).

2. The equipment factors in the transfer regression equation (although specifically selected for their potential importance to the task under investigation) accounted for only a small proportion of the total variance.

If these patterns hold, then two important practical conclusions can be drawn:

1. Fewer data will be required to establish a transfer performance surface, since a lower-order model is indicated. (A sequential strategy, described later in this paper, ensures the most effective and economical approach to fit the appropriate model to the data.)

2. When prediction of transfer is critical, more rather than fewer factors should be investigated during the screening phase, including task and training factors as well as additional equipment variables not considered in this study.

---

*Mosteller and Tukey (1977, Chapter 2) write: "The word indication is a vague concept intended to include, at one extreme, all of the classical descriptive statistics... but also, at another extreme, to include any hints and suggestions that might prove informative to a reasonable man... What indication is not is inference or treatment of uncertainty..." (pp. 25-27).

Economy as a Function of Purpose

The purpose for which a study is performed can markedly affect the amount of data collection required, independent of the number of factors being investigated. If the study is done primarily to identify the important factors, fewer data will be required than if its purpose is to say with reasonable confidence that slight differences are in fact not real differences between levels. In the latter case, the power of the test requires many more observations to accept the null hypothesis with reasonable confidence.

Similarly, fewer data are ordinarily required in experiments intended to identify critical factors rather than to write an equation representing a response surface. Generally, the former can be satisfied with a modified fractional factorial design; the latter requires that the data collection process continue until the model being created fits the experimental data. The present design represents a compromise between these two purposes. By adding the centerpoints, lack of fit and curvilinearity can be evaluated, yet each factor cannot be identified individually.

Of particular interest to transfer of training experiments, where economy is a still more serious consideration, is the distinction between an experiment intended only to identify which factor affects transfer of training the most and one that must provide a measure of transfer. In Table 13, the coefficients of each factor (with the exception of Transfer Vehicle Configuration) are identical for the transfer performance scores and the transfer effect scores in which the mean performance of each control group has been subtracted from the transfer performance score of the corresponding transfer group. Thus the mean difference between levels for the individual equipment variables is the same whether or not a control group is used.

From the standpoint of design, this means that we can predict which equipment combination will yield the highest performance on the transfer configuration without collecting independent control data. In the present study, this would have represented a savings of 30%. Roscoe (1980, Chapters 16 and 17) has discussed the various ways of measuring the transfer and cost effectiveness of simulator training. However, in transfer of training studies for simulator design purposes, the raw performance scores are sufficient during the early "screening" phase, in which the purpose is simply to identify the best configuration. It is in the evaluation phase later in a program that experiments should be performed with control groups.

Predicting Transfer from Training Performance Data

If it were possible to eliminate the transfer phase of a training study considerable savings would be elicited. This could be done if it were found that a significant relationship exists between training performance scores and transfer. The observed correlation of .37 between the 56 training performance scores and the transfer effect scores, with 54 degrees of freedom, would be expected to occur only about five times in 1000 by chance. Nevertheless, an *r* of .37 accounts for less than 14 perceht of the variability exhibited among the transfer effect scores and therefore has relatively little practical prediction value.

Applying Principles of Transfer to Achieve Economy

Some psychologists believe they can predict the transfer potential of simulator configurations on the basis of principles that have been proposed at one time or another regarding the necessary relationships between training and transfer (see Roscoe, 1980, Chapters 15-22). If this were in fact possible this could eliminate or at least reduce the data collection effort required to make decisions regarding the design of the training simulator. However, at present, these principles are at best imprecise generalizations, largely a product of unstructured empirical observation and tenuous analogy with findings from abstract laboratory experiments on verbal and motor learning. Although the present study was not designed to isolate individual principles of transfer, some information regarding so-called transfer principles was obtained.

The training performance principle. It has been suggested that the size of a transfer of training experiment might be reduced by eliminating those factors during transfer that fail to influence performance critically during the training phase. This principle may be stated as follows: *Factors that fail to influence performance during training will not critically influence performance during transfer*. Since in holistic experiments involving ten or more factors it is not uncommon to find many noncritical effects, were this theory a valid and sufficient one, considerable savings could be elicited.

But there are reasons to suspect the validity of this principle. Simon (1971) has discussed how task difficulty is a hidden variable that can confound the magnitude of the effects of the manipulated variables. By way of illustration for certain situations, if a task is too easy, no differences may be observed in the performance obtained at two levels of a factor. As the task becomes more difficult, differences in performance at the two levels will begin to appear. When the task becomes too difficult, performances at the two levels will become the same again.

In transfer, quite frequently the overall operational task will be more complex and difficult than the simulated task due to adverse environmental factors not represented in the simulation. Carryover effects, another name for transfer, can be confounded with task difficulty level in the same way. If this is the case, we cannot say without a great deal of additional knowledge whether effects will or won't be larger or smaller between training and transfer.

If one examines the experimental results (Tables 12 and 13), it can be observed that between the training and transfer phases some large effects become small and some small effects become large. It has already been noted that the total transfer surface, except as affected by the transfer vehicle configurations, was relatively flat. Still, an inspection of the data can be enlightening. In Table 18 are listed all statistically significant interaction coefficients for transfer performance from Table 13 and the corresponding training performance coefficients from Table 12.

TABLE 18. INTERACTION TERMS THAT WERE STATISTICALLY SIGNIFICANT (p < .05) IN TRANSFER PHASE*

	Coefficients and Chance Probability				
	Training	p	Transfer	p	Ratio
DL x TT	6.2	.644	35.7	.040	5.8
TM x PT	19.2	.155	-39.1	.026	2.0
TM x XL	-29.9	.065	-58.2	.006	1.9
TM x XQ	5.9	.509	25.5	.029	4.3
CG x TT	-9.6	.477	-46.7	.009	4.9
CG x XQ	10.4	.250	24.7	.034	2.4

*From Tables 12 and 13

There are six interaction effects that were not larger than might have been expected by chance during training but were during transfer. The coefficients in these cases increased from two to six times in magnitude. While this study, involving a relatively simple perceptual-motor task, may not be representative of the carrier landing task or a simulator's visual "fidelity" factors, it does provide an empirical indication of the danger of accepting the theory that what isn't important during training will not be important for transfer.

Lincoln (1978) performed a transfer of training experiment involving four longitudinal compensatory tracking tasks that differed with regard to system dynamics. Two factors at two levels each were investigated: short period natural frequency and damping characteristics. In discussing the results of the study, he wrote:

> "The most important result of this study concerns the differential effects of the two major dynamic variables. One of them, damping, greatly influenced task difficulty but was of little importance in determining the effects of transfer of training. The second variable, natural frequency of the system affected performance in exactly the reverse manner. Its influence was relatively unimportant with regard to task difficulty, but it appeared to be of primary importance in determining the amount of transfer of training that occurred" (p. 88).

The last statement is another empirical example wherein the theory proposed at the beginning of this section was contradicted.*

---

*In support of the holistic approach to equipment design research, the following quotation was taken from the discussion in Lincoln's paper: "Unfortunately for the designer, as Muckler, Obermayer, Hanlon, and Serio (1961) have shown, the complex nature of transfer effects and the need to recognize the possiblity that other variables may interact with frequency and damping, makes broad generalizations dangerous. In their second report, these investigators found that control gain settings could drastically alter the patterns of transfer that were observed. It appears, therefore, that designers of manual control systems are presently faced with problems of considerable complexity with only limited empirical information on which to base their design decisions. A continuing effort to untangle the interrelated effects of training and transfer would seem to be appropriate" (p. 89).

The similarity principle. One of the most frequently stated training and transfer generalizations is that associated with the similarity between training and transfer conditions, commonly referred to as "fidelity" when applied to simulators. This principle asserts:

> Transfer of training from simulator to aircraft is a positive function of the degree to which the simulator faithfully reflects the characteristics of the aircraft.

The effect of this principle has been the design and development of simulators that have a high degree of physical similarity to the real airplane, a costly decision of unknown payoff. Still, it is generally recognized (but sometimes forgotten) that task similarity does not necessarily depend on a faithful physical representation of reality. What it does depend on necessarily is a faithful representation of the responses that must be learned, and the conditions needed to elicit those responses.

The chief problem with similarity (or fidelity) in a complex system in a real-world environment is that it is not always easy to define or measure. In fact, as Simon (1979, Section VI) has noted, it is a multivariate concept that differs in form and meaning for different physical components of a complex simulation. Then too, only certain components are critical in simulating particular tasks, and as yet there are no adequate principles for deciding which are and which aren't critical. The concept of similarity is further complicated by the interaction between stimulus similarity and response similarity and their differential effects on positive and negative transfer.

To make matters worse, these are not the only principles that have an effect on training and transfer, and others, such as adaptive feedback, may override the effects of similarity in any situation. Until these principles have been adequately dimensionalized and empirically evaluated together and in context, they will continue to offer only superficial aid in the design of complex simulators that are optimized both for pilot training and for cost of ownership, maintenance, and operation.

From the results of the present experiment, two dimensions of task similarity could be examined, factor fidelity, as represented by the distance between levels of the various experimental factors, and relative difficulty, as represented by the different training and transfer vehicle configurations. Stated in its negative form, the fidelity principle asserts that: The farther apart the training and transfer levels of a particular factor are, the lower the transfer effect. The relative difficulty principle asserts that: More positive transfer will be elicited when task difficulty shifts from hard to easy than from easy to hard.

The transformations used to change coded experimental coefficients of each equipment/training factor to "fidelity" and "difficulty" coefficients are shown in Table 19. With these "similarity" coefficients the experimental design is no longer orthogonal. Still a stepwise regression analysis could be and was performed to see which of the six fidelity (factor-distance) and six difficulty (factor-direction) variables most influenced transfer performance. With an F of 4.00 to enter and exit the equation, only three of the 12 terms appeared when transfer performance scores were analyzed and only two terms when transfer effect scores were the criterion.

TABLE 19. TRANSFORMATIONS USED TO CHANGE CODED EXPERIMENTAL COEFFICIENTS TO SIMILARITY COEFFICIENTS FOR EACH FACTOR AND CONDITION

Coded Experimental Coefficients		Coded Similarity Coefficients			
		Fidelity (Distance)		Difficulty (Direction)	
Training Code	Transfer Code	Actual	Code	Actual	Code
-1 (Hard)*	-1 (Hard)	0	-1	No Change	0
-1	0 (Central)	1	0	Hard to Easy	-1
-1	+1 (Easy)	2	+1	Hard to Easy	-1
+1 (Easy)	-1	2	+1	Easy to Hard	+1
+1	0	1	0	Easy to Hard	+1
+1	+1	0	-1	No Change	0
0 (Central)	-1	1	0	Easy to Hard	+1
0	0	0	-1	No Change	0
0	+1	1	0	Hard to Easy	-1

*These are the preexperimental assumptions regarding the relative difficulty of performing at this level.

Table 20 lists the terms that appeared, their coefficients in the equations, the incremental proportion of variance each contributed, the standard regression coefficients of each, and such summary statistics as the F-value, the standard error of estimate, the standard error of coefficients, and the adjusted R^2.

TABLE 20. SUMMARY OF STEPWISE REGRESSION ANALYSES OF INTERVENING FIDELITY AND DIFFICULTY FACTORS WITH AN F OF 4.00 TO ENTER AND EXIT

Source	Coefficients	Proportion of Variance	Standard Coefficients
Transfer Performance Scores			
Intercept	1688.0		
Direction-TM	+157.0	.23	.34
Direction-CG	+157.0	.14	.34
Direction-CO	+106.0	.05	.23
			Adjusted R^2 = .38
F = 12.44 (3/52), p < .001			
Standard Error of Estimate = 278			
Standard Error of Coefficients = 51.7			
Transfer Effects			
Intercept	-11.3		
Distance-TM	+61.9	.11	.33
Direction-TM	+54.0	.07	.26
			Adjusted R^2 = .15
F = 5.73 (2/53), p < .01			
Standard Error of Estimate = 145			
Standard Error of Coefficients = 25.5			

When training performance scores were introduced as an independent variable along with the 12 similarity variables and a stepwise regression analysis was performed for transfer performance and transfer effect scores, the results shown in Table 21 were obtained:

TABLE 21. SUMMARY OF STEPWISE REGRESSION ANALYSES OF INTERVENING FIDELITY AND DIFFICULTY FACTORS WITH TRAINING PERFORMANCE SCORES AS AN INDEPENDENT VARIABLE AND AN F OF 4.00 TO ENTER AND EXIT

Transfer Performance Scores

Source	Coefficients	Proportion of Variance	Standard Coefficients
Intercept	874.9		
Training Performance	+0.475	.09	.30

Adjusted R^2 = .07

F = 5.16 (1/54), p < .0
Standard Error of Estimate = 341
Standard Error of Coefficients = .209

Transfer Effects

Source	Coefficients	Proportion of Variance	Standard Coefficients
Intercept	-429.1		
Training Performance	+0.243	.14	.34
Direction-TM	+57.9	.08	.27
Distance-TM	+51.1	.07	.28

Adjusted R^2 = .25

F = 7.06 (3/52), p < .0
Standard Error of Estimate = 136.4
Standard Error of Training Performance Coefficient = 0.085
Standard Error of Coefficients = 23.0

Several results important for simulator design are indicated by the above statistics:

1. The positive coefficients of the similarity variables indicate that the results agreed with the principles stated earlier. For those factors that had an effect, when the factor level between training and transfer changed from hard to easy there was a greater reduction in RMS error than when it changed from easy to hard. Also, the shorter the distance between the two levels, the greater the reduction in RMS error.

2. In this study, the direction of change in difficulty had stronger effects than the distance of change in factor levels.

3. The adjusted R^2 values showed that the similarity variables did not improve our prediction of transfer scores over that based on training performance scores. In fact, there was an interactive effect depending on whether training performance or similarity variables were used to predict transfer performance or transfer effect scores. None of the combinations did as well as the original equipment/training variables.

No generalizations regarding the usefulness of similarity variables as intervening factors can be drawn here. Before one can discard these approaches, however, the following limitations of this analysis must be considered: (a) It was an adjunct effort, an afterthought, and the study was not designed to obtain this kind of information. (b) The performance and similarity variables in this study were intercorrelated, which means that these solutions are not unique. (c) The measures of similarity used here were unsophisticated. (d) The similarity variables are only two of the possible classes that could have been included. (e) No interaction effects among the similarity variables were included in these analyses.

The primary importance of this exercise lies in the fact that it illustrates what might be done were a comprehensive set of intervening factors developed and a primary investigation performed. The information obtained from such a study would help resolve the question of the contribution of intervening factors to predictions between simulators and test vehicles.

Fidelity and Quasi-Transfer Experiments. The costs of conducting simulator-to-air transfer studies have severely limited that type of research. Then too, certain difficult-to-accept risks exist when a pilot must be tested in the air after having been trained on a less than optimum simulator training configuration. For these reasons, quasi-transfer experiments, in which both training and transfer occur in simulators, can fill an important gap. The problem with quasi-transfer studies is that their results can be questioned on the grounds that the observed transfer was to a simulator, not to an airplane.

Unless the investigator can show that performance in the transfer configuration of the simulator does not differ in any critical way from performance in the airplane there always will be room for doubt regarding the information a quasi-transfer experiment provides. For this reason, if transfer of training research is anticipated and if economy is an issue, then some secondary effort must be devoted to establishing the relationship between performance in the most physically faithful configuration of the simulator and performance in the aircraft. It is also important to develop multidimensional similarity principles that will permit the transfer relationships between simulator and aircraft to be defined quantitatively.

Alternative Economical Designs

In the carrier landing performance study carried out earlier at the VTRS laboratory (Westra, et al., 1981), 128 experimental conditions were used to evaluate the effects of ten equipment and environmental factors on carrier landing. The likelihood of running a transfer of training experiment of that size is meager, both because of the extended costs of training and the difficulty of finding the large number of subjects that would be needed.

Ten factors can still be investigated, however, using only 32 experimental conditions. One limitation of such a design is that 32 conditions are close to the minimum that would provide an acceptable level of precision. Each mean is based on 16 observations, which is a reasonable number. However, with highly variable data, our confidence limits will be quite wide. Still the design can pick out those factors that have an important effect on performance but may miss the marginal ones. Another limitation of a design of that size is that, while it isolates all main effects from each other and from all two-factor interactions, it does not isolate all of the two-factor interaction effects from one another. The two-factor interactions are aliased in strings.

If 32 experimental conditions are too costly, there are still more economical designs that might be employed. It is important, however, to emphasize that one uses these designs to obtain the best empirical information possible under extremely limited conditions, and that the best may or may not be suitable for certain purposes. The decision to use such designs is not an experimental question. It is one that management, informed of the tradeoffs involved, must make.

For example, there are main-effect designs by Plackett and Burman (1946) (see Simon, 1973) that increase in steps of four rather than powers of two. Thus Plackett-Burman designs are based on 4, 8, 12, 16, 20, 24, 28, 32 conditions rather than 4, 8, 16, 32 conditions of the usual fractional factorial plans. With ten factors, for example, one can measure all main effects using only 12 experimental conditions. However, these effects will be partially confounded with two-factor interactions ($r = \pm 0.333$) and with three-factor interactions ($r = \pm 0.354$).* Furthermore, each interaction will be found in several strings, each partially confounded with a number of other effects. In the standard 2^{k-p} designs, confounding--if it occurs--is total ($r = \pm 1.0$) and unique to a particular set of other effects.

---

*Dr. David Weinman, Hollins College, Virginia provided the information regarding the intercorrelations for the Plackett-Burman designs.

With these Plackett-Burman designs, by adding 12 more experimental conditions -- the "foldover" (Simon, 1973) of the original plan -- the main effects can now not only be isolated from one another but also from all two-factor interactions. The two-factor interactions remain partially correlated with other two-factor interactions in several different strings, but the cost has been reduced from 32 to 24 conditions. Only the main effects remain correlated with the three-factor interactions ($\underline{r} = \pm 0.333$).

Thus for certain numbers of factors, Plackett-Burman designs enable smaller studies to be done than the 2^{k-p} designs. In the above example, there was a 25% savings in data collection. The precision with which the means can be measured dropped 13% with the smaller design. The higher-order effects become more difficult to untangle than when fully aliased. On the other hand, being partially correlated, no single large interaction effect is likely to bias the main effect estimates seriously. With a correlation of 0.333, the overlap is approximately 11%. These designs might be considered when it is essential that the study be kept small, when less precision can be accepted regarding the main effects, and when there is little likelihood that there will be a subsequent need to isolate interaction effects.

Using a Sequential Strategy

In the present study, a fixed design was employed. That is, the size and form of the design, its resolution and other characteristics were selected before the experiment began. This is not the way to perform holistic experiments economically. Instead, only the data needed to fit the lowest-order surface should be collected initially, and the model tested to see how well it fits the empirical data. Then if the fit is poor, more data would be collected to fit the next higher-order surface. This process would continue until the fit is adequate. In the present study, the fixed design was used because of concern that the time and dollar limitations might prematurely terminate a sequential effort. Consequently, a plan was selected that would guarantee at least a Resolution V design.

What the present study suggests regarding the application of a sequential approach to transfer of training is that we should build increasingly complex models on the basis of transfer scores rather than training scores. This is indicated, at least in the current study, because the transfer surface is less complex than the training surface and therefore requires fewer data points. Then too, transfer scores are the values of ultimate interest. Only additional effort will determine to what extent the results of the present experiment can be generalized, but for the present, the approach employed is appropriate.

SECTION VII

CONCLUSIONS AND DISCUSSION

This study demonstrated an efficient and economical approach for collecting multifactor, multicriterion transfer of training data. The approach is particularly useful in the early stages of a simulator design program, such as VTRS, when many alternatives should be considered and the individual contributions of component design variables should be evaluated separately from overall simulator effectiveness.

The unique nature of this study should be noted. It was the first transfer of training experiment:

1. To examine as many as six equipment/training factors in a single study.

2. To examine a broad spectrum of training configurations at one time. Forty-nine configurations were examined here.

3. To train only a single subject on each of the 48 main training vehicle configurations, reducing the number required without sacrificing precision.

4. To use more than one transfer vehicle configuration in a single experiment. Three were used here.

5. To provide data in equation form that would answer specific questions regarding transfer from various training simulator configurations not directly investigated.

The cost of this study in data-collection time was approximately one-eighth of what it might have cost to study the effects of the six factors and the three transfer configurations in a series of two-factor studies to obtain all main and two-factor interaction effects. The precision of this study was higher and potential bias less.

Multifactor data collection plans of this type are most effective in the early phases of a simulator design program before specific configurations have been selected. Conventional transfer of training research designs are better fitted for use at the end of a design program when a few configurations have been selected and the objective is to quantify the transfer effectiveness of each.

The application of this data collection plan provides a number of practical features seldom obtained from conventional transfer of training experiments or rational analyses by design engineers and psychologists. For example, in the early stages of a simulator design program, the actual features of the transfer criterion vehicle and consequently

the simulator requirements may not be firm. This data collection plan provides transfer data across a broad spectrum of conditions (training and test) so that when the airplane features are eventually selected, relevant transfer data will be available.

Isolating the effects of potentially critical variables in the simulator provides better transfer data with which to make engineering decisions than do gross measurements of total simulator effectiveness. It enables the designer to identify negative contributions of specific components that might otherwise be hidden by positive overall results. The data collection plan, by providing multifactor data in equation form, not only allows estimates to be made of the effectiveness of simulator configurations not investigated in the study, but also provides an overview that enables tradeoffs to be made more precisely between system performance and system costs.

The following results in this study have direct applications for future transfer-of-training efforts provided subsequent investigations demonstrate their generality:

1. The transfer performance surface is approximated by a lower-order model than that required for the training performance surface and should be used as the criterion for collecting more data if a sequential data collection plan is employed to obtain maximum economy.

2. The correlation between training and transfer performance for different design configurations was positive but too low for practical predictive purposes.

3. Effects that are strong in training may not be strong in transfer and vice versa.

4. Intervening similarity factors combined with performance scores may increase predictability, although insufficient experimental data exist to isolate and evaluate intervening predictive factors at this time.

5. Where time, pilot availability, and dollars are extremely limited, marginal experimental plans can provide empirical data regarding transfer effectiveness of simulator design variables, but with increased risk of error.

6. Additional economy may be achieved by eliminating control groups in the early phases of a simulation design program when the purpose is to select those combinations of factor levels that produce the highest transfer possible rather than measure a particular configuration's effectiveness.

REFERENCES

Campbell, D. T., and Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1963.

Collyer, S. C., and Chambers, W. S. AWAVS, a research facility for defining flight trainer visual requirements. Proceedings of the Human Factors Society 22nd Annual Meeting. Santa Monica, CA: Human Factors Society, 1978.

Connor, W. S. and Young, S. Fractional factorial designs for experiments with factors at two or three levels. Washington, DC: U.S. Government Printing Office, National Bureau of Standards, Applied Mathematics Series 58, 1961.

Draper, N. R. and Hunter, W. G. Transformations: Some examples revisited. Technometrics, 1969, 11, 23-40.

Lincoln, R. S. Transfer of training on manual control systems differing in short-period frequency and damping characteristics. Human Factors, 1978, 20, 83-89.

Lindquist, E. F. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin, 1953.

Mosteller, F., and Tukey, J. W. Data analysis and regression. Reading, MA: Addison-Wesley, 1977.

Muckler, R. A., Obermayer, R. W., Hanlon, W. H., Serio, F. R., and Rockway, M. R. Transfer of training with simulated aircraft dynamics: (II) Variations in control gain and phugoid characteristics. Wright-Patterson AFB, OH: WADD Technical Report 60-615, 1961.

North Atlantic Treaty Organization, Advisory Group for Aerospace Research and Development. Fidelity of Simulation for Pilot Training. London: Technical Editing and Reproduction, Ltd. AGARD Advisory Report No. 159, 1980.

Plackett, R. L. and Burman, J. P. The design of optimum multifactorial experiments. Biometrika, 1946, 33, 305-325.

Roscoe, S. N. Aviation psychology. Ames, IA: The Iowa State University Press, 1980.

Simon, C. W. Considerations for the proper design and interpretation of human factors engineering experiments. Culver City, CA: Hughes Aircraft Company, Technical Report No. P73-325, 1971.

Simon, C. W. Economical multifactor designs for human factors engineering experiments. Culver City, CA: Hughes Aircraft Company, Technical Report No. P73-326A, 1973. (AD A035-108)

Simon, C. W. *Design, analysis, and interpretation of screening designs for human factors engineering research*. Westlake Village, CA: Canyon Research Group, Inc., Technical Report No. CWS-03-77A, 1977. (AD 056-985)

Simon, C. W. *Applications of advanced experimental methodologies to AWAVS training research*. Orlando, FL: Naval Training Equipment Center, Technical Report NAVTRAEQUIPCEN 77-C-0065-1, 1979. (AD A064-332)

Simon, C. W. *Applications of advanced experimental methods to visual technology research simulator studies: Supplemental techniques*. Orlando, FL: Naval Training Equipment Center, Technical Report NAVTRAEQUIPCEN 78-C-0060-3, 1981. (AD A095-633)

Westra, D. P., Simon, C. W., Collyer, S. C., and Chambers, W. S. Investigation of simulator design features for carrier landing tasks. *Proceedings of the IMAGE Generation/Display Conference II*. Williams AFB, AZ: AFHRL Operations Training Division, 1981.

Williams, A. C., Jr. and Flexman, R. E. *An evaluation of the Link SNJ operational trainer as an aid in contact flight training*. Port Washington, NY: Office of Naval Research, Special Devices Center, Contract N6ori-71, Task Order XVI, TR 71-16-5.

APPENDIX A

TRAINING VEHICLE CONFIGURATIONS IN THE FOUR EXPERIMENTAL GROUPS EXPRESSED IN CODED VALUES CORRESPONDING TO REAL-WORLD VALUES SHOWN IN TABLE 1

TABLE A1

UNIQUE TRAINING VEHICLE CONFIGURATIONS (COMBINATIONS OF + AND - FACTOR LEVELS) FOR THE 16 INDIVIDUAL PARTICIPANTS IN GROUP A WHO TRANSFERRED TO THE HARD TRANSFER VEHICLE CONFIGURATION

GROUP A	Training Vehicle Configuration					Training Trials
Participant	CO	DL	TM	PT	CG	TT
01	-	-	-	-	-	-
02	+	+	+	+	+	+
03	+	-	-	-	+	+
04	-	+	+	+	-	-
05	-	+	-	-	+	+
06	+	-	+	+	-	-
07	+	+	-	-	-	-
08	-	-	+	+	+	+
09	-	-	+	-	+	-
10	+	+	-	+	-	+
11	+	-	+	-	-	+
12	-	+	-	+	+	-
13	-	+	+	-	-	+
14	+	-	-	+	+	-
15	+	+	+	-	+	-
16	-	-	-	+	-	+

TABLE A2

UNIQUE TRAINING VEHICLE CONFIGURATIONS (COMBINATIONS OF + AND - FACTOR LEVELS) FOR THE 16 INDIVIDUAL PARTICIPANTS IN GROUP B WHO TRANSFERRED TO THE CENTRAL TRANSFER VEHICLE CONFIGURATION

GROUP B	Training Vehicle Configuration					Training Trials
Participant	CO	DL	TM	PT	CG	TT
17	-	-	-	+	-	-
18	+	+	+	-	+	+
19	+	-	-	+	+	+
20	-	+	+	-	-	-
21	-	+	-	+	+	+
22	+	-	+	-	-	-
23	+	+	-	+	-	-
24	-	-	+	-	+	+
25	-	-	+	+	+	-
26	+	+	-	-	-	+
27	+	-	+	+	-	+
28	-	+	-	-	+	-
29	-	+	+	+	-	+
30	+	-	-	-	+	-
31	+	+	+	+	+	-
32	-	-	-	-	-	+

TABLE A3

UNIQUE TRAINING VEHICLE CONFIGURATIONS (COMBINATIONS OF + AND - FACTOR LEVELS) FOR THE 16 INDIVIDUAL PARTICIPANTS IN GROUP C WHO TRANSFERRED TO THE EASY TRANSFER VEHICLE CONFIGURATION

GROUP C	Training Vehicle Configuration					Training Trials
Participant	CO	DL	TM	PT	CG	TT
33	+	-	-	-	-	+
34	-	+	+	+	+	-
35	-	-	-	-	+	-
36	+	+	+	+	-	+
37	+	+	-	-	+	-
38	-	-	+	+	-	+
39	-	+	-	-	-	+
40	+	-	+	+	+	-
41	+	-	+	-	+	+
42	-	+	-	+	-	-
43	-	-	+	-	-	-
44	+	+	-	+	+	+
45	+	+	+	-	-	-
46	-	-	-	+	+	+
47	-	+	+	-	+	+
48	+	-	-	+	-	-

TABLE A4

CENTRAL TRAINING VEHICLE CONFIGURATION
FOR THE EIGHT INDIVIDUALS IN GROUP D WHO TRANSFERRED
TO THE CENTRAL TRANSFER VEHICLE CONFIGURATION

GROUP D	Training Vehicle Configuration					Training Trials
Participant	CO	DL	TM	PT	CG	TT
49	0	0	0	0	0	0
50	0	0	0	0	0	0
51	0	0	0	0	0	0
52	0	0	0	0	0	0
53	0	0	0	0	0	0
54	0	0	0	0	0	0
55	0	0	0	0	0	0
56	0	0	0	0	0	0

APPENDIX B

GENERAL INSTRUCTIONS READ ALOUD TO ALL PARTICIPANTS WHILE THEY SILENTLY FOLLOWED ON A COPY

"These are the instructions for the task you will be asked to perform. I will read the instructions aloud while you follow along silently."

"We are conducting an experiment to compare various types of flight training displays. We know that you are unfamiliar with this type of task but you will find that you do better after a while. We are only interested in which of several displays is the best for training."

TASK

"The task is a simplified version of what a pilot performs while flying an airplane. Centered on the display in front of you are two symbols; a box (□) which is the target, and a figure (—⏀—) that looks something like a rear view of an airplane. By moving the stick left and right, you will be able to control the bank angle of the airplane symbol which in turn controls its position relative to the target. Because we are concerned with teaching pilots to steer a plane, only left-right movements of the stick will affect the simulated airplane you are flying. Your task will be to keep the airplane close to the target as much of the time as you can. We will measure how closely you are able to follow the target and tell you your score at the end of each one-minute trial."

VARIATION OF TASKS

"Different participants will be asked to perform on different types of displays. You may be asked to perform on displays in which only the target symbol will move left and right. The airplane symbol will rotate but will always remain in the center of the screen. In other types of displays, both the airplane symbol and the target symbol will move to show their relative positions. In either case, you will be able to control the position of the 'actual' airplane relative to the target by always banking or turning toward the moving target symbol. The 'circle' that usually forms the 'body' of the airplane can move out ahead of the cross-shaped part of the airplane, in which case it predicts where the airplane will be a moment later. Moderate use of the control stick will prevent over-control of the airplane."

Following joint reading of instructions while participant remains in control chair, experimenter flies one demonstration trial, saying only the following:

(1) "Now I'm going to demonstrate your task."
(2) "I am trying to keep the airplane as close to the target as I can."
(3) "Notice that to do this I always keep turning the airplane symbol toward the target by banking in that direction."

If questions are asked, they are to be answered as directly and simply as possible. Then the instructions are jointly read a second time beginning with the section labeled TASK.

PROCEDURE

"You will fly a series of trials that are approximately one minute in length. At the end of each trial, the display will momentarily go blank and then reset to the starting position with all the symbols in the center of the screen. You will be told your score after each trial. Your score is based on your average distance from the target. The lower the score, the better the performance. Before each trial, I will signal you with the words READY, BEGIN. The information we obtain here will be important in designing pilot training systems. All we ask is that you do your best during the entire session. Please notify me if the display does not operate like I said it would."

"Are there any questions?"

"Will you please explain what your task is and how you are to accomplish it?"

(The participant replies accordingly.)

"If there are no further questions, wait for my READY, BEGIN signal."

The participant will now fly the three matching trials before which the experimenter will say,

"You will now fly a series of trials with an experimental display."

Next, the participant will fly the two masking trials and the block of experimental trials before which the experimenter will say,

"You will now fly a series of trials with a different experimental display."

"READY, BEGIN"

APPENDIX C

PREANALYSIS OF EXPERIMENTAL DATA

Characteristics of data often determine how they should best be analyzed. Experimental designs alone do not guarantee high quality information. When economical multifactor designs are employed, the opportunities for finding effects due to chance or unintentional biases makes preanalysis central to the experimental approach. The following are some of the more important results of a preanalysis of the data from this experiment.

DATA CHARACTERISTICS

Relevant characteristics of distributions of experimental data samples include questions of normality, homogeneity of variances, randomness of sampling, and sources of experimental bias.

Transformation

Performance was originally measured in root mean square error per trial. The question is: Should we analyze the data as is or should they first be transformed? There are numerous criteria to apply before deciding to work with transformed data; one rule of thumb is that when the deviations of minimum and maximum scores from the mean differ by more than two to one, some form of transformation is in order.

The ratios both before and after a logarithmic transformation are shown in Table C1, and it appears that (a) a transformation of some kind was desirable and (b) the log transformation served its purpose. As a methodological note, the choice of the logarithmic transformation was made rationally because it has been found to normalize RMS error scores and equalize their variances in the past. The preferred approach, however, would be to let the data determine the choice of transformation, using a method proposed by Draper and Hunter (1969; Simon, 1981) for handling multivariate transformations.

Central Tendency Measure

Inspection of the raw data revealed that individual performances varied considerably from trial to trial. It was decided to combine the data over several trials in a single representative score. The best index of central tendency of a number of trials was desired. The question is: Should this measure be the mean or median of each individual's set of trials? The means and standard deviations of the mean and median scores of all 56 participants in the experimental groups for these sets are given in Table C2. Correlations between mean and median scores for the 56 participants in the experimental groups were: training performances, first five trials, 0.990, last five trials, 0.996; and transfer performances, first five trials, 0.994.

TABLE C1

COMPARISON OF DISTRIBUTIONS OF RAW AND TRANSFORMED PERFORMANCE SCORES FOR THE 56 PARTICIPANTS IN GROUPS A, B, C, AND D COMBINED

RMS ERROR	TRAINING PHASE		TRANSFER PHASE	
Smallest	15.96		12.05	
Difference		-35.44		-36.70
Mean	51.40		48.75	
Difference		+136.96		+423.31
Largest	188.36		472.06	
Ratio of Absolute Differences		3.86		11.53
1000 LOG RMS				
Smallest	1203		1081	
Difference		-508		-607
Mean	1711		1688	
Difference		+564		+986
Largest	2275		2674	
Ratio of Absolute Differences		1.11		1.62

TABLE C2

MEANS AND STANDARD DEVIATIONS OF THE MEANS AND MEDIANS OF SCORES FOR BLOCKS OF FIVE TRIALS EACH FOR THE 56 PARTICIPANTS IN GROUPS A, B, C, AND D COMBINED

	First 5 Training Trials		Second 5 Training Trials		First 5 Transfer Trials	
	Mean	Median	Mean	Median	Mean	Median
Mean	1782	1775	1718	1711	1697	1688
Standard Deviation	246	242	237	232	372	373

It could be argued that either the mean or median of a set of trials would be an acceptable measure. However, because sizable discrepant individual scores occasionally occur within a set of trials, it was decided that using median values for each individual would provide the best protection against the presence of occasional outliers. If there were no greatly discrepant values, means and medians would be essentially the same. If there were occasional ones for some participants, the medians would provide the better indication of central tendency. Thus all individual scores are the medians of sets of trials.

Group Matching

The experimental design consisted of seven blocks of experimental conditions. Three blocks are those for the three experimental groups associated with the three transfer configurations. One block was at the center of the experimental space. The remaining three represent the three control groups. The question is: Did these seven groups differ initially in ability on this task? The means and standard deviations of each group's initial matching scores are shown in Table C3.

A test of the differences among the seven means, taking group size into consideration, yielded an F-ratio with 6 and 73 degrees of freedom of 1.43, a value that might be expected to occur by chance 21 times in 100. Bartlett's test of the homogeneity of variance among the seven groups yielded an approximate F of 1.43 and a p of 0.20. On such evidence one cannot reject the hypothesis that the groups were randomly drawn from a homogeneous population.

TABLE C3

MEANS AND STANDARD DEVIATIONS OF MATCHING SCORES FOR EACH GROUP

	Experimental Groups (N = 16 each)			Center Group (N = 8)	Control Groups (N = 8 each)		
	A	B	C	D	E	F	G
Mean	1704	1697	1662	1664	1750	1663	1688
Standard Deviation	98	80	86	36	68	60	92

Bias Check

The data had been collected by three experimenters over a period of several weeks and at different times of the day or evening. Did the different experimenters, time of day, or phase of the data collection effort introduce any detectable biases in the data? Correlations among those three factors and several of the key performance measures are shown for the combined data in Table C4. There is no evidence that any bias was introduced by these variables.

CORRESPONDENCE WITH EXPECTATIONS

One means of examining the quality of empirical data is to see whether they correspond with prior expectations. Applications of economical multifactor data collection plans require extensive pre-experimental analysis to anticipate subsequent results based on whatever theory and fact can be brought to bear on the problem. Explaining results after the data have been collected provides little confidence in their interpretation and leads to the traditional sequence of experiments to verify the results of previous experiments. Gross disagreements between expectations and data warn the investigator that there is something wrong either with the theory or the data.

In either case, discrepancies serve as focal points for a more extensive examination of data, whereas correspondence between expectations and results is definitely encouraging. Nevertheless, correspondence remains a source of contentment only to the extent that the data conform to several criteria. With multifactor designs, however, when complex patterns of data are observed, the confidence produced from conforming data is much higher than would be the case with simple designs in which the data can only increase, decrease, or reverse expectations along a single continuum of conditions. In this study, a limited preanalysis enabled only a few expectations to be evaluated.

TABLE C4

CORRELATIONS RELATING POTENTIAL SOURCES OF EXPERIMENTAL BIAS TO DIFFERENT PERFORMANCE SCORES (N = 56)

	Potential Sources of Bias		
Scores	Experimenters	Experiment Phase	Time of Day
Matching Score	+.017	-.133	-.003
Training Performance	-.049	+.028	-.003
Transfer Performance	-.127	-.029	-.097
Transfer Effect	-.151	+.028	-.094

Relative Difficulty

In selecting the factor levels, pretests were made to confirm expectations as to which levels would lead to better and poorer performances. A coded -1 value was assigned to the level yielding the poorer performance. Similarly, estimates were made regarding the relative difficulty of the three transfer configurations, coded -1, 0, and +1. Other less design-oriented predictions were made regarding possible interactions among factors, but these preanalyses were scanty and unsystematic. In retrospect, our inability to predict factor interactions during training and transfer was due considerably to the fact that even experienced investigators have had little or no opportunity to see the effects of complex relationships and interactions demanded for preanalyses of this type.

Let us see, however, to what extent the data do or do not behave as expected. For example: Are the three transfer configurations relatively hard, intermediate, and easy as predicted? Median performances for the three control groups on the Hard, Central, and Easy system configurations, respectively, were: 2179, 1597, and 1368, with lower scores being better. The order was as expected, although the Central configuration was not midway between Hard and Easy.

Amount of Learning

Since this was a transfer of training study, it is appropriate to ask: To what extent did the participants learn? Figure C-1 shows the learning curves created by connecting the means of scores from the matching, masking, first-five training, last-five training,* first-five transfer and last-five transfer trials for all 56 participants, and additional curves for each of the four groups individually. These curves behave remarkably well, and in accordance with expectations. However, they are averages and therefore sanitized. During training the individual curves varied considerably, both within and among themselves.

The mean proportion of variance accounted for by the residuals for all 56 participants after that due to linear, quadratic, and cubic trends was isolated is 0.62; the range ran from .95 to .13. The regularity of these statistics among the three groups, A, B, and C suggest that the equipment configurations rather than individuals were contributing to this within learning curve variability. Furthermore, not only were the learning curves somewhat erratic, but also there was considerable variability between curves when different configurations were used. Some exhibited a reasonable amount of learning and some showed none. Still the correlations among the main factor effects and the linear performance effects during training were all low (i.e., median, .13; range .11 to .28; N = 56) showing that the factors individually did not influence how much learning took place to any significant degree.**

Additional analyses were performed on the data from the primary experimental groups. For each participant the median performance of the last five training trials (6-10 or 25-30, depending on the amount of training) was compared with that from the first five training trials. For each group, the difference between each individual's scores from those two sets of data was divided by the median performance value of all first-five training trials to obtain an indication of each individual's percent improvement. Only three of the 24 participants who received 10 training trials showed an improvement of 10% or better whereas 14 of the 24 who received 30 training trials improved 10% or better. Four improved more than 20% and three more than 30%.

---

*Last five training trials may have been the 6th to 10th, 16th to 20th, or 26th to 30th, depending on the group.

**Note: These correlations reflect the degree of linear relationship that existed. If a curvilinear relationship existed, it might not be detected and would in fact result in a low correlation measure of the type used here. No plots of the data were made.

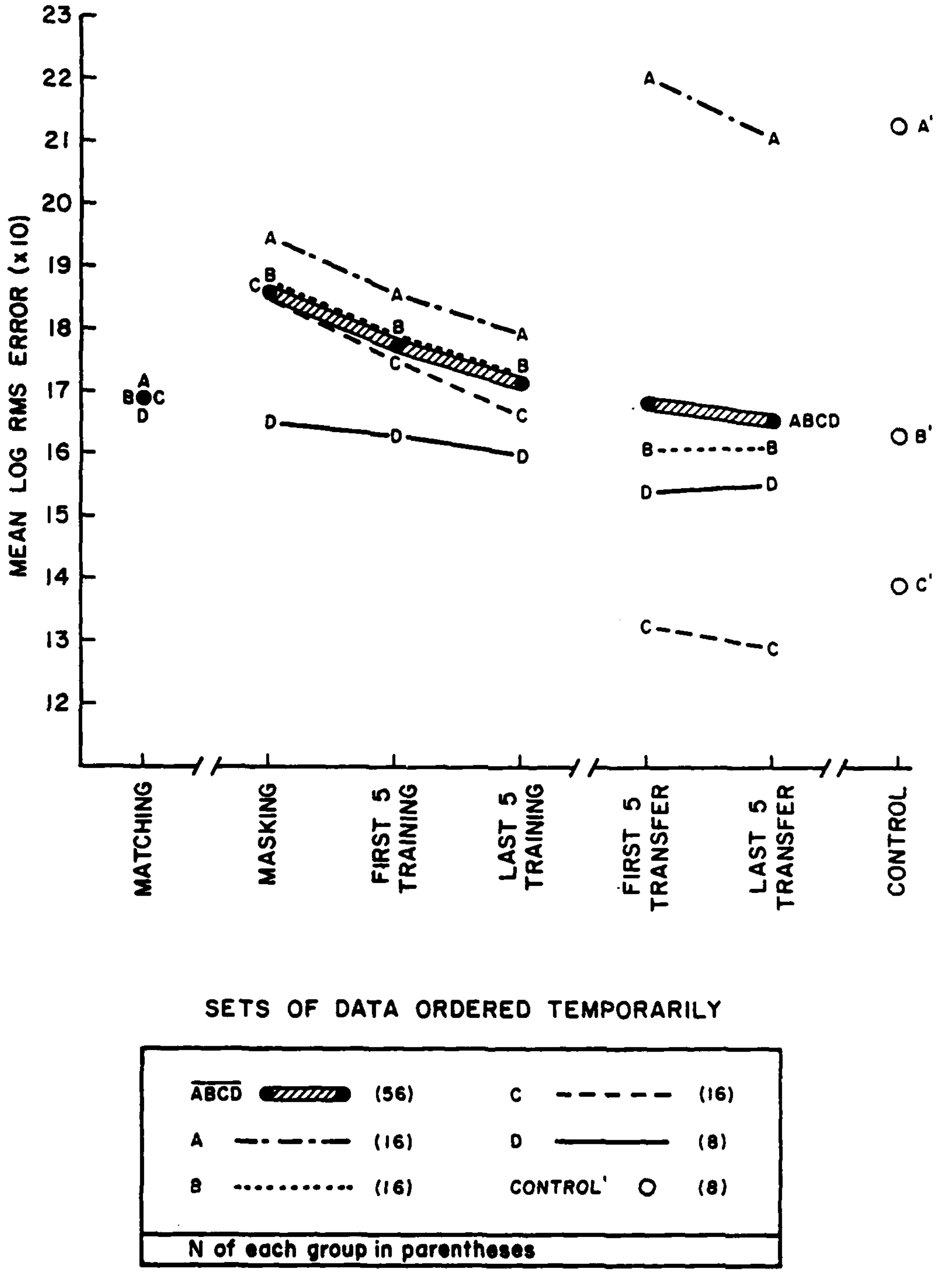


Figure C1. Mean training and transfer curves for Groups A, B, C, and D singly and combined.

These results suggest that the performance surface is shallow. While a great deal of pretransfer learning is not an absolute prerequisite for a transfer study, lack of it suggests that the task was either too easy or too difficult. This does not negate the study's usefulness as a demonstration of the effectiveness of the multifactor approach to transfer of training research. It may limit the interpretation of the factor effects in a particular study, because in the real world, task difficulty levels in a simulator need to match those in the counterpart airplane if results of comparative studies in the simulator are to be trusted.

APPENDIX D

PERCENT OF TRANSFER

The transfer effect score used in this study is not commonly employed by psychologists to measure transfer of training (see Roscoe, 1980, Chapter 16). Here, only differences between performance scores on the transfer vehicles by trained and control participants were used in the main analyses, whereas percent transfer scores are normally used to compare devices. This latter measure is obtained by dividing the transfer effect score by the control group's score and multiplying by 100, a simple monotonic transformation.

Because of the nature of the task in this study, either transfer effect scores or percent transfer scores might have been used to demonstrate the practical advantages of the holistic approach to transfer of training research. However, because there were three transfer vehicle configurations rather than one, the set of 56 percent transfer scores was not the result of a simple transformation of the transfer effect scores; the relationship was more complex.

For example, if a transfer effect score in Group A were 100 and another in Group C were also 100, the two percent transfer scores would not be equal since the divisor for Group A is larger than the divisor for Group C. This difference would not change the interpretation of any equipment or training factor (since each is orthogonal to the transfer vehicle factor), nor their interactions. However, the relationships between the experimental factor and percent transfer scores across the three different transfer vehicles could be different from those between the experimental factors and transfer effect scores.

The transfer effect scores and percent transfer scores correlate .99, which for all practical purposes implies identity. This suggests that it would not matter which measure is used, at least for the results obtained in this experiment. The percent transfer scores and the training performance scores for the last five trials correlate .41. This compares with the correlation of .37 between the transfer effect scores and the same training performance. The difference is trivial.

Table D1 for the analysis of percent transfer scores parallels Table 13 for both transfer effect and transfer performance scores. Two facts stand out when the tables are compared. First, the direction of transfer (whether positive or negative) is the same for every source of variance. Second, seven effects in each table with chance probabilities of less than .05 were the same. The ten largest standard regression coefficients were associated with the same sources of variance in both tables. Although some differences in order occurred, the results are quite similar overall.

TABLE D1: ANALYSIS OF TRANSFER PERFORMANCE EFFECTS: GROUPS A, B, C, AND D

Source of Effect	Regression Coefficient	Standard Error	Standardized Coefficient	*p*
INTERCEPT	-0.017			
Control Order	-0.011	0.009	-.12	.229
Display Lag	-0.020	0.009	-.22	.032
Tracking Mode	0.014	0.009	.16	.117
Prediction Time	0.006	0.009	.06	.525
Control Gain	0.017	0.009	.18	.067
Training Trials	0.001	0.009	.01	.904
TVC, Linear (XL)	-0.039	0.011	-.35	.001
TVC, Quadratic (XQ)	0.021	0.007	.28	.007
CO x DL	0.002	0.011	.02	.856
CO x TM	-0.014	0.009	-.14	.165
CO x PT	0.014	0.009	.15	.160
CO x CG	0.010	0.009	.11	.292
CO x TT	0.005	0.009	.05	.596
CO x XL	-0.017	0.011	-.15	.131
CO x XQ	0.005	0.006	.08	.410
DL x TM	-0.012	0.009	-.13	.202
DL x PT	0.156	0.009	.16	.131
DL x CG	0.003	0.009	.03	.748
DL x TT	0.022	0.009	.23	.031
DL x XL	-0.009	0.011	-.08	.407
DL x XQ	0.004	0.006	.06	.536
TM x PT	-0.017	0.009	-.19	.080
TM x CG	-0.001	0.011	-.01	.905
TM x TT	0.008	0.009	.09	.387
TM x XL	-0.028	0.011	-.25	.016
TM x XQ	0.013	0.006	.20	.048
PT x CG	0.014	0.009	.15	.147
PT x TT	-0.011	0.011	-.12	.323
PT x XL	0.013	0.011	.11	.261
PT x XQ	-0.008	0.006	-.12	.207
CG x TT	-0.024	0.009	-.26	.017
CG x XL	-0.005	0.011	-.05	.617
CG x XQ	0.014	0.006	.21	.035
TT x XL	0.006	0.011	.05	.585
TT x XQ	0.000	0.006	.00	.998

A stepwise regression analysis of the percent transfer data brought the same four factors into the equation as found in the transfer effects analysis. These were, in order of importance, XL (.12), XQ (.08), DL x TT (.08), and TM x XL (.06), with the proportion of variance each accounted for shown in parentheses. The adjusted multiple R for this percent transfer equation was .29, compared to .30 for the equation based on transfer effects.

In summary, although the findings from the analyses of transfer effects and percent transfer scores were for all practical purposes identical, this need not be valid in general. For this reason the scores analyzed in each case should be the ones of greatest interest to the user.

DISTRIBUTION LIST

Naval Training Equipment Center 134
Orlando, Florida 32813

Commander 3
HQ, TRADOC
Attn: ATTNG-PA
Ft. Monroe, Virginia 23651

Center Library 3
Naval Personnel Research and
Development Center
San Diego, California 92152

Dr. Ralph R. Canter 1
U.S. Army Research Institute
Field Unit
P. O. Box 16117
Fort Harrison, Indiana 46216

Defense Technical 12
Information Center
Cameron Station
Alexandria, Virginia 22314

PERI-OU 1
U.S. Army Research Institute for the
Behavioral & Social Sciences
5001 Eisenhower Avenue
Alexandria, Virginia 22333

OASD (MRA & L)/Training 2
Room 3B922, Pentagon
Washington, D.C. 20301

Dr. Ralph Dusek 2
U.S. Army Research Institute for
the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, Virginia 22333

The Van Evera Library 1
Human Resources Research Organization
300 North Washington Street
Alexandria, Virginia 22314

Library 1
Division of Public Documents
Government Printing Office
Washington, D.C. 20402

Director 2
Training Analysis & Evaluation Group
Department of the Navy
Orlando, Florida 32813

HumRRO/Western Division/Carmel Office 1
27857 Berwick Drive
Carmel, California 93923

U.S. Coast Guard (G-P-1/62) 1
400 Seventh Street, SW
Washington, D.C. 20590

Personnel & Training Research Programs 3
Office of Naval Research
(Code 458)
Psychological Sciences Div.
800 N. Quincy Street
Arlington, Virginia 22217

National Aviation Facilities 1
Experimental Center
Library
Atlantic City, New Jersey 08405

AFHRL/TSZ 2
Brooks AFB, Texas 78235

DISTRIBUTION LIST (cont'd)

American Psychological Assoc. 1
Psyc. INFO Document Control Unit
1200 Seventeenth Street, NW
Washington, D.C. 20036

AFHRL Technology Office 1
Attn: MAJ Duncan L. Dieterly
NASA-Ames Research Center
MS 239-2
Moffett Field, California 94035

Center for Naval Analyses 1
Attn: Dr. R. F. Lockman
2000 N. Beauregard Street
Alexandria, Virginia 22311

Dr. J. Huddleston 1
Head of Personnel Psychology
Army Personnel Research Establishment
c/o RAE, Farnborough
Hants, ENGLAND

OUSDR&E (R&AT) (E&LS) 1
CDR Paul R. Chatelier
Room 3D129, The Pentagon
Washington, D.C. 20301

Dr. Jesse Orlansky 1
Science and Technology Division
Institute for Defense Analyses
400 Army-Navy Drive
Arlington, Virginia 22202

Chief of Naval Operations 1
OP-987H
Attn: Dr. R. G. Smith
Washington, D.C. 20350

Scientific Technical Information Office 1
NASA
Washington, D.C. 20546

AFOSR/NL (Dr. A. R. Fregley) 1
Bolling AFB
Washington, D.C. 20332

Human Factors Society 2
Attn: Bulletin Editor
P. O. Box 1369
Santa Monica, California 90406

National Defense University 1
Research Directorate
Ft. McNair, D.C. 20319

Commanding Officer 1
Air Force Office of Scientific Research
Technical Library
Washington, D.C. 20301

Dr. D. G. Pearce 1
Behavioral Sciences Division
Defense and Civil Institute of Environmental Medicine
P. O. Box 2000
Downsview, Ontario M3M, CANADA

Technical Library 1
OUSDR&E
Room 30122
Washington, D.C. 20301

Commander 2
Naval Air Systems Command
AIR 340F
Attn: CDR C. Hutchins
Washington, D.C. 20361

Chief 1
ARI Field Unit
P. O. Box 476
Ft. Rucker, Alabama 36362

DISTRIBUTION LIST (cont'd)

Chief of Naval Operations 1
OP-115
Attn: M. K. Malehorn
Washington, D.C. 20350

Technical Library 1
Naval Training Equipment Center
Orlando, Florida 32813

Chief of Naval Operations 1
OP-596
Washington, D.C. 20350

Commander 1
Naval Air Test Center
CT 176
Patuxent River, Maryland 20670

Office of Deputy Chief of Naval Operations 1
Manpower, Personnel and Training (OP-01)
Washington, D.C. 20350

Assistant Secretary of the Navy 1
Research, Engineering & Systems
Washington, D.C. 20350

HQ Marine Corps 1
Code APC
Attn: LTC J. W. Biermas
Washington, D.C. 20380

Chief of Naval Operations 1
OP-593B
Washington, D.C. 20350

Scientific Advisor 1
Headquarters U.S. Marine Corps
Washington, D.C. 20380

Dr. Martin Tolcott 1
Office of Naval Research
800 N. Quincy Street
Department of the Navy
Arlington, Virginia 22217

Commander 1
Naval Air Development Center
Attn: Technical Library
Warminster, Pennsylvania 18974

Naval Research Laboratory 1
Attn: Library
Washington, D.C. 20375

Chief of Naval Education and Training Liaison Office 6
AFHRL/OTLN
Williams AFB, Arizona 85224

Dr. Donald W. Connolly 1
Research Psychologist
Federal Aviation Administration
FAA NAFEC ANA-230 Bldg. 3
Atlantic City, New Jersey 08405

Chief of Naval Material 1
MAT 08D2
CP5, Room 678
Attn: Arnold I. Rubinstein
Washington, D.C. 20360

Commanding Officer 1
Naval Education Training Program and Development Center
Attn: Technical Library
Pensacola, Florida 32509

Commander 1
Naval Air Systems Command
Technical Library
AIR-950D
Washington, D.C. 20361

Chief of Naval Education and Training 1
Code 01A
Pensacola, Florida 32509

DISTRIBUTION LIST (cont'd)

Commander 1
Pacific Missile Test Center
Point Mugu, California 93042

Commander 1
Naval Air Systems Command
AIR 413
Washington, D.C. 20361

Commanding Officer 1
Naval Aerospace Medical Research Laboratory
Code L5
Department of Psychology
Pensacola, Florida 32512

Dr. Thomas Longridge 1
AFHRL/OTR
Williams AFB, Arizona 85224

Dr. Kenneth Boff 1
ARAMRL/HEA
Wright Patterson AFB, Ohio 45433

CAPT James Goodson 1
Code L-32
Naval Aerospace Medical Research Laboratory
Pensacola, Florida 32512

Dr. Genevieve Haddad
AFOSR/NL
Bolling AFB, D.C. 20332

Dr. David C. Nagel 1
LM-239-3
NASA Ames Research Center
Moffett Field, California 94035

Federal Aviation Administration 1
Technical Library
Bureau Research & Development
Washington, D.C. 20590

Commander 1
Naval Weapons Center
Human Factors Branch (Code 3194)
Attn: Mr. Ronald A. Erickson
China Lake, California 93555

Mr. Robert Wright 1
Aeromechanics Lab
(USAAVRADCOM)
Ames Research Ctr, MS 239-2
Moffett Field, California 94035

Lt Col Jefferson Koonce 1
USAFA/DFBL
USAF Academy, Colorado 80840

CDR Joseph Funaro 1
Code 602
Human Factors Engineering Division
Naval Air Development Center
Warminster, Pennsylvania 18974

Dr. Will Bickley 1
USARI Field Unit
P.O. Box 476
Fort Rucker, Alabama 36362

LMED

-8