AD-A252 371

ARI Research Note 92-39

(1)

Feasibility Study for Predicting Human Reliability Growth Through Training and Practice

John C. Lowry, Virginia A. Rappold, and Michael M. Copenhaver

Allen Corporation of America

Manned Systems Group
John L. Miles, Jr., Chief

MANPRINT Division
Robin L. Keesee, Director

May 1992

ARMY RESEARCH INSTITUTE

United States Army
Research Institute for the Behavioral and Social Sciences

Approved for public release; distribution is unlimited.

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

---

Research accomplished under contract
for the Department of the Army

Allen Corporation of America

Technical review by

Marshall A. Narva

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
1992, May

3. REPORT TYPE AND DATES COVERED
Final Dec 89 - Dec 90

4. TITLE AND SUBTITLE
Feasibility Study for Predicting Human Reliability Growth Through Training and Practice

5. FUNDING NUMBERS
DAHC35-89-D-0027
63007A
793
1206
C02

6. AUTHOR(S)
Lowry, John C.; Rappold, Virginia A.; and Copenhaver, Michael M.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Allen Corporation of America
209 Madison Street
Alexandria, VA 22314

8. PERFORMING ORGANIZATION REPORT NUMBER
--

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and Social Sciences
ATTN: PERI-S
5001 Eisenhower Avenue
Alexandria, VA 22312-5600

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARI Research Note 92-39

11. SUPPLEMENTARY NOTES
Contracting Officer's Representative, Charlie Holman

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release;
distribution is unlimited.

12b. DISTRIBUTION CODE
--

13. ABSTRACT (Maximum 200 words)

This report examines the feasibility of developing a stand alone, quantitative Human Reliability Growth Model (HRGM) that predicts the impact of training variables on soldier performance. Such a model would incorporate learning curve fitting techniques to predict the impact of training variables on performance and would be based on empirical data from the behavioral and social science literature and available government data bases.

This report describes the effort to collect empirical data on the effects of learning and practice on human performance. In addition, the report contains a review of the theoretical literature involving human learning and practice in which the nature and application of learning curves and curve fitting techniques are derived and summarized.

The results of this effort reveal that, out of approximately 3,000 research titles and abstracts reviewed, only 27 articles meet minimal criteria for use in developing the HRGM. It was concluded that, although a theoretical basis for developing an HRGM exists, the data could not support its development.

14. SUBJECT TERMS
Human reliability growth
Human performance modeling
Training and performance
Human learning and practice
Learning curve fitting techniques
Training and performance literature review

15. NUMBER OF PAGES
210

16. PRICE CODE
--

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) is interested in developing a Human Reliability Growth Model (HRGM) that accurately predicts relationships between relevant training variables and soldier task performance. Of specific interest would be a model that implements data related to the initial training of soldiers at the U.S. Army Training and Doctrine Command (TRADOC) schools and the effect of training time on soldier performance of military tasks.

Approaches to this problem are typified by methods that are highly qualitative in nature (e.g., anecdotal observations). However, these approaches result in no direct, quantitative evaluation of the impact of training on soldier task performance. An approach that combines the benefits of existing qualitative data collection techniques with the precision of quantitative data would be most useful to the Army.

This report examines the feasibility of developing an HRGM that predicts the impact of training variables on soldier performance. Such a model would be configured to be both stand alone (i.e., requiring no additional analytic support for the performance of analyses using the model) and quantitative. In order to satisfy these criteria, the HRGM would have to be based on empirical data obtained from the behavioral and social science literature, as well as available government data bases.

Thus, the primary focus of this feasibility study was to examine whether empirical data could be collected to support the development of an HRGM. As an additional part of the conceptualization of the HRGM, research underlying human learning curves and curve fitting techniques was reviewed to predict the impact of training on soldier performance based on learning curve relationships.

Results of this study were used to evaluate the feasibility of developing a stand alone HRGM based on quantitative relationships between numerous training variables and soldier performance. Moreover, results from this study may serve as a basis for future models or algorithms requiring such literature.

FEASIBILITY STUDY FOR PREDICTING HUMAN RELIABILITY GROWTH THROUGH TRAINING AND PRACTICE

CONTENTS

LIST OF TABLES

LIST OF FIGURES

FEASIBILITY STUDY FOR PREDICTING HUMAN RELIABILITY GROWTH THROUGH TRAINING AND PRACTICE

Introduction

The Army Research Institute (ARI) requires the development of a Human Reliability Growth Model (HRGM) to predict the relationship of numerous training variables on soldier task performance. Of particular interest is a focus on (a) initial training of soldiers as would occur at TRADOC (U.S. Army Training and Doctrine Command) schools, (b) effect of training time (time to learn) on soldier performance of military tasks, and (c) empirically derived relationships between training and performance. Such a model should be configured to be both stand alone (i.e., requiring no additional analytic support for the performance of analyses using the model) and quantitative. The development of such a model has to be based on empirical data obtained from the behavioral and social science literature and available government data bases. Thus, the purpose of this project was to examine the feasibility of developing this model given the risks attendant to collecting the sort of empirical data (i.e., the data may not be available or may not exist) required to successfully develop such a model.

The literature on learning curve fitting techniques was reviewed as part of the feasibility study. Incorporating current views and research underlying human learning curves, a model was conceptualized based on relationships defined through learning curve fitting techniques. Specifically, the Growth Model would predict the effect of training on performance based on consideration of human (i.e., soldier) capabilities and potential improvements in performance due to training. The characteristics of a HRGM would be

- Provides the capability to predict the impact of training variables (e.g., time, media, and method) on soldier performance based on learning curve relationships

- Provides formalized quantitative relationships of the impact of training on human performance based on empirical data obtained from actuarial data sources

- Makes use of taxa-based (i.e., task-based) relationships to order and organize the impact of training on performance.

This feasibility study became an effort to address the following question as a primary focus of the research effort:

Do sufficient amounts of suitable data exist?

In order to address this question, two parallel activities were pursued. The first was an effort to collect all the available literature that presents empirical data on the effects of human learning and practice on human performance. This activity consisted of (1) literature searches of on-line data bases of human behavioral and social science literature abstracts, (2) personal contacts with government points-of-contact for sources of government owned data bases containing the required empirical data, (3) 'brute-force' investigation of personal and corporate research libraries to identify relevant literature, and (4) limited direct contact with current researchers known to have conducted research in these areas. The research collected through this manner was reviewed, archived, abstracted, and logged in a data base for later use. Out of approximately 3,000 research titles and abstracts reviewed, 27 research articles were selected as meeting minimal criteria for acceptable utility.

The second activity was a comprehensive review of the theoretical literature involving human learning and practice, with particular focus on human learning curve fitting techniques. This review encompassed 70 years of research beginning with Thurstone (1919), in which the nature and application of learning curves were derived and summarized. The review also helped in providing focus to the search and collection of empirical learning and performance data. A final step was the development of criteria for determining the acceptability of empirical data for use in developing an HRGM that meets the project criteria for being stand alone and quantitative.

The final activity in this project was to evaluate the data collected with respect to the criteria established for their acceptance. This analysis was used to determine the feasibility of developing a successful growth model, given project constraints.

This report contains a technical section in addition to the introduction. The first part of the technical section discusses the theoretical bases for the measurement and prediction of human learning and practice and performance. The second part presents a detailed description of the empirical data collection effort and a summary of the results obtained. The third part presents a discussion of the overall feasibility of developing a successful HRGM that meets project requirements.

This report also contains a section of five appendixes. Appendix A presents the empirical data and research abstracts from the 27 articles that met minimal criteria for developing the HRGM, and Appendix B contains references for the 27 articles.

Appendix C is a bibliography of all articles reviewed for inclusion in the HRGM. Appendix D presents detailed algebraic derivations of various learning curve functions reported in the literature. Finally, Appendix E contains a bibliography of supplemental research that is relevant to the project, but not central to the effort.

Theoretical Bases for the Measurement and Prediction of Human Learning and Practice

The relation between training variables and soldier performance must be determined in order to develop a stand alone HRGM. The most pervasive approach used by learning researchers to assess the relationship between training (i.e., learning and practice) and performance (e.g., time or trials) is curve fitting. Essentially, curve fitting describes changes in performance over time or trials using the typical "acquisition" curve prevalent in studies of skill learning and retention (Lane, 1986b).

The purpose of this chapter is to review the literature on the measurement and prediction of the major aspects of training and performance. First, this encompasses the behavioral science literature involving the learning curve and its relation to the measurement and prediction of the effects of human learning and practice on the attainment of skilled performance. Second, the behavioral science literature on the measurement and prediction of the effects of training variables on the performance of military tasks are reviewed. Finally, the requirements for a HRGM and associated data requirements are identified.

Theoretical Roots of the Learning Curve and Its Heuristics

Since Thurstone (1919) laid out his formal definition of the learning curve; the "amount of attainment gained per unit of practice decreases as practice increases", the 'ubiquitous'[1] learning curve has come to dominate the area of human learning and practice as the metric of choice among behavioral scientists. While not the first to develop such constructs, Thurstone derived the prototypical equation for fitting cumulative errors (or conversely cumulative successes) with a hyperbolic function or curve (Lane, 1986b).[2] He identified the notable characteristics of the learning curve which over time have more or less held constant across areas of human learning research. In addition to the description of the general appearance of the learning curve noted above, the more notable characteristics among those Thurstone identified include:

---

[1]The word 'ubiquitous' has been employed by a number of more recent learning researchers (e.g., Newell & Rosenbloom, 1981) as a description of the persistence of the learning curve as the metric of choice and because few other operational learning constructs have evolved as alternatives.

[2] For the sake of simplicity and to avoid the ongoing debate as to the optimum form of the learning curve function, the major learning curve equations (as well as their derivations) that have been discussed in the literature over the years, are available in Appendix D of this report.

- Learning curves sometimes take the form of a "positive acceleration at the initial stage of learning, plateaus during the course of learning, and erratic advancement of attainment."

- "The use of an equation for this relation" (i.e., the attainment of skills over trials or practice time) "enables one to predict the limit of practice before it has been attained, provided that the learning follows the law of diminishing returns."

- Rate of learning can be differentiated from the limit of practice (i.e., asymptote) since these are "undoubtably independent" metrics.

- The learning curve equation can "be of service in the analysis of the relation between the variability in learning and other mental attributes."

- "All questions of transfer of training may be investigated by the learning equation and the transfer effect may be differentiated into psychological components."

In 1930, Thurstone revisited his earlier effort, strengthened the premises he postulated by adding new inferences to them, and evolved more complex forms of his learning curve equations. These additional characteristics of the learning curve include:

- "The two principal variables in learning are: (a) practice, and (b) attainment. Attainment" is "an increasing function of practice. Practice may be measured either in terms of repetition or in terms of time devoted to it, and effort is assumed to be constant. Repetitions may be counted as successes and failures."

- "Learning consists in a series of separable acts, that some of these are counted as successful and that others are counted as errors. Each act whether right or wrong will be counted as a unit of practice."

- "Attainment" (also expressible in various ways) can be expressed as a measure of "the probability that an act will be counted as successful. This measure of attainment is proportional to and consequently for our purposes synonymous with the number of successful acts per unit time."

In a pair of articles, Thurstone (1919; 1930) identified the major learning curve parameters and the typical behavior (or form) that the learning curve equations often take. These parameters and the behavior (form) of the learning curve have dominated research up to the present time. Thurstone identifies:

- the major learning curve parameters:

  (a) rate of learning
  (b) asymptote ('limit of practice')
  (c) initial learning rate ('positive acceleration at the initial stage of learning'),

- the hyperbolic form of the learning curve equation,

- the rules for the measurement and prediction of learning and practice effects on skill attainment, including:

  (a) assumptions underlying the methodology for the measurement of practice,
  (b) assumptions underlying the measurement of successes or failures to demonstrate an operation, and
  (c) approximate equivalence of 'the probability that an act will be counted as successful' and 'the number of successful acts per unit time', and

- the practical implications of measuring and predicting human learning and practice effects on human performance:

  (a) transfer of training and transfer effects
  (b) relation of human learning and practice to other human mental attributes.

Each of these characteristics (and others added by later researchers) will be discussed in a section for determining the requirements for a HRGM and the data requirements associated with such a model.

Families of Learning Curves

Since Thurstone (1919; 1930), several additional functions or families of learning curves have been identified based on regularities in the shapes of the curves.

Power law. One of the most prevalent learning curves identified is the power law of practice or power function (Lane, 1986b; Newell & Rosenbloom, 1981). The power function fits a wide variety of skill data such as perceptual-motor skills, motor skills, and cognitive skills, and is "ubiquitous when one measures the log of performance time against the log of trial number" (Newell & Rosenbloom, 1981, p. 2). Its shape is negatively accelerated. That is, the gain in performance on a trial decreases as practice increases (see Figure 1, Curve 1-A).[3]

The power function takes the form:

$$T = A + B\ (N + E)^{-R}$$

where

T = time to respond or to perform a task (completion of one unit of output on a given trial),

A = asymptote or highest level of performance attainable,

B = performance or output on first trial,

N = number of trials or time units (index of practice),

E = prior learning (transfer from prior experience or learning required to attain entry level performance),

R = rate variable (amount of change in Y or T with one unit change in N or average slope of the curve).

In power functions, the independent and dependent variables are usually transformed logarithmically, and this log-log transformation yields a relatively straight line (Lane, 1986b; Newell & Rosenbloom, 1981).

Hyperbolic. The hyperbolic function (introduced above during the discussion of Thurstone's research) is derived from the power function but requires the fitting of one less parameter to the data because it assumes an implied rate variable (R) of -1 (Lane, 1986b).

[3] It should be noted that the families of curves depicted in Figure 1 are generalized to show their differences.

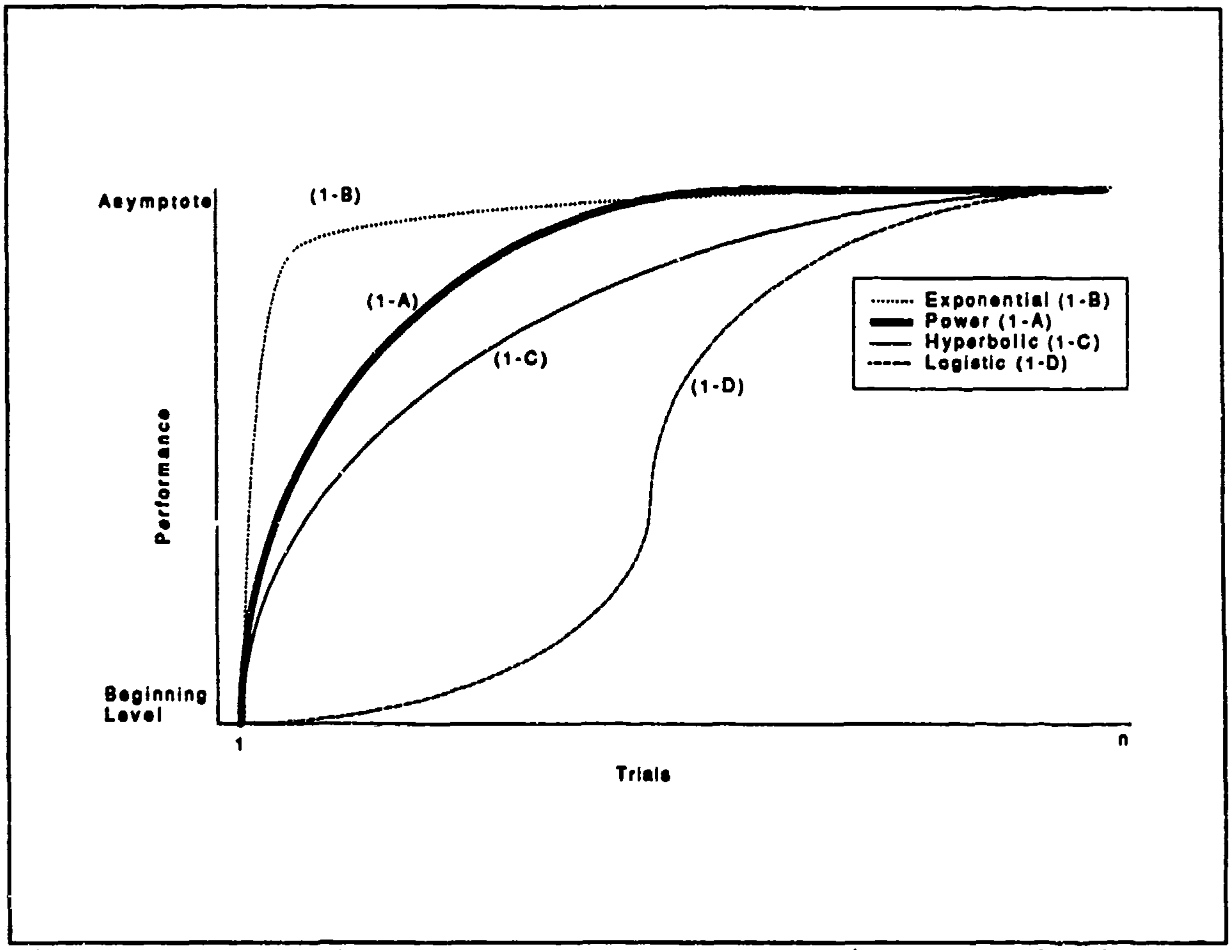


Figure 1. General shapes of power, exponential, hyperbolic, and logistic curves.

The hyperbolic function is:

$$T = A + B / (N + E)$$

where

T = time to respond or perform a task,

A = asymptote

B = performance or output on first trial

N = number of trials

E = prior learning.

When B is positive, the function will be decreasing, whereas the function will be increasing when B is negative. A generalized hyperbolic curve is shown in Figure 1, Curve 1-C.

Exponential. The third family of curves found in learning data is the exponential function. The exponential function differs from the power function in that it is semi-logarithmic; only one variable is transformed logarithmically. In addition, the exponential curve decreases and increases in a more rapid fashion than the power curve because the amount learned on each trial does not decrease as a function of N (the number of trials). Thus, the exponential curve "produces a curve of constant acceleration, producing a much 'steeper' curve than the power law" (Lane, 1986b, p. 30). An example of a generalized exponential curve is shown in Figure 1, Curve 1-B.

The form of the exponential function is:

$$T = A + B\,e^{-RN}$$

where

T = time to respond or perform a task,

A = asymptote

B = performance or output on first trial

e = natural logarithm

R = rate variable

N = number of trials

Logistic. The final member in the family of curves is the logistic function. The logistic curve is not as common in learning data as power or exponential curves, and appears to be derived primarily from data involving performance ratings by instructors in training or assessment situations (e.g., check pilot evaluations performed in the aviation training community prior to, during, and following formal training of student pilots). The logistic function is an "S" shaped or sigmoid curve that is both positively and negatively accelerated. In other words, the curve rises slowly in early trials, accelerates rapidly in the middle trials, and levels off in later trials (see Figure 1, Curve 1D). As seen in Figure 2, the logistic curve may contain two inflection points (i.e., the point in a curve where positive acceleration changes to negative acceleration or where negative acceleration changes to positive acceleration) compared to power and exponential curves that contain a single inflection point (Lane, 1986b).

The logistic function takes the form:

$$Y = A / [1 + (B - A) e^{-kN}]$$

where

Y = performance score, ratings, or any other measure that is represented by increasing values

A = asymptote

B = performance or output on first trial

e = natural logarithm

k = an implicit function of R, representing the percentage of learning to be accomplished on each trial; $k = R / [Y (A - Y)]$

N = number of trials

The logistic function is often referred to as the "autocatalytic equation" because previous learning "catalyzes" subsequent learning. That is, previous learning speeds up subsequent learning resulting in a "snowballing" effect (Lane, 1986b; Spears, 1983).

This discussion has sought to focus on the major topics and parameters identified in the literature on human learning curves and their mathematical forms. We have sought to avoid the controversy regarding the "correct" form of the learning curve function extant in the available literature. Rather, we believe that it is possible that a number of these curves are on approximately equal footing with respect to the legitimacy of their use for any given task and training situation. As a consequence, the selection of a learning curve function should be based on empirical data specific to the training context and nature of the task. Later researchers have identified some of the parameters, conditions, and constraints on the use of learning and performance data. This body of research is the focus of the next section of this chapter.

Applications of Learning Curves to Military Training

Recently, Spears (1983) has used learning curve theory to explain transfer of training (TOT). Although Spears explains TOT within the context of training devices (particularly flight training using simulators), his principles are generalizable to a large variety of training situations and tasks. According to Spears, the transfer of skills from training (such as on a training device) to the operational environment (such as to

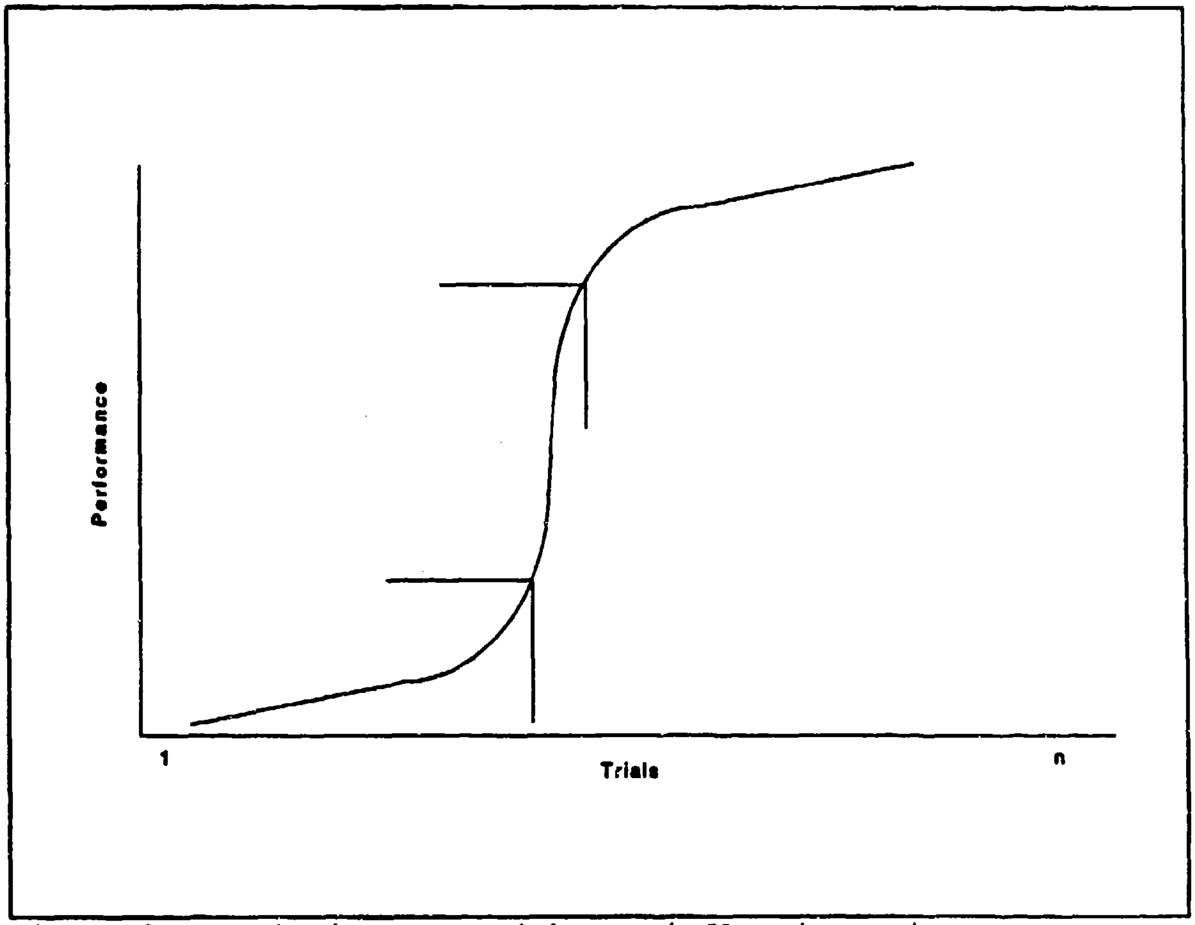


Figure 2. Logistic curve with two inflection points.

operational equipment) is essential in training. Therefore, it is important to recognize and understand the empirical indicators or parameters of learning curves on which researchers, such as Spears, focus during the measurement and evaluation of training.

Spears describes four basic parameters of learning curves that are important for the study of learning and transfer in applied settings:

a) beginning level,
b) asymptote,
c) rate of learning, and
d) inflection point.

The first parameter, beginning level, refers to the level of performance on a task before any practice has occurred. According to Spears (1983), examining the beginning level of curves may help in comparisons of different methods of training. For example, comparing the beginning levels of different methods of training and transfer may help to determine which method would be more cost effective. Spears cautions, however, that beginning level performance is notoriously unstable when applied to the context of predicting later learning based on initial learning. For this reason, he limits the application of beginning level as a parameter and cautions researchers to collect data on a sufficient number of trials to avoid the instability.

The second parameter, asymptote, is the highest level of performance attainable for all practical purposes.[4] It indicates what is or could be accomplished during training. Asymptote is important as an indicator of the point at which a given stage of skill integration is completed and designates the level of achievement at that stage. Asymptote may also be useful when comparing skill integration between two methods of training (i.e., one method requires less training to reach asymptote than another). It is most useful for comparisons of training time, devices, and methods.

The third parameter, rate of learning, is the measure used to assess training effectiveness (the change in achievement per unit of change in the independent variable). It is a primary parameter for enabling decisions regarding training resource allocations (e.g., time, materials, and people) and amount of practice.

The last parameter, inflection point, is the point in a curve where positive acceleration changes to negative acceleration or vice-versa. Examining the inflection point(s) of learning curves may be important when examining the integration of multi-component skills; they could reveal when integrative processes are complete. Inflection points could also indicate when a level of proficiency is reached in a transfer task. However, there have been no practical applications of using inflection points to study learning and transfer to date.

Clearly, any of these four basic parameters are potentially useful to address specified training concerns. They should be considered for potential application within the context of a HRGM.

Other researchers, notably Lane (1986b), have used learning curve theory to describe the general influence of key variables on military training and education. Initially, Lane attempted to develop a generalized learning curve which was to be used to estimate the duration of training. However, he found that the published literature would not support the development of a

---

[4]However, attainment of skilled performance can continue to a significant degree many thousands of practices or trials beyond what many researchers would designate as asymptote.

quantitative prediction of training time. He cites the following as explanation for the failure to develop a generalized learning curve:

- the literature on learning curves did not contain the type of quantitative data needed to estimate training time,
- key information was often omitted in published studies,
- many of the tasks studied were too simple,
- many tasks were not representative of military tasks,
- many of the curves reported in the literature were group curves (based on group data), and
- curves based on individual performance data, when reported, differed from the group curves.

Therefore, Lane (1986b) dropped the effort to develop a generalized learning curve and instead focused his efforts on the description of conditions affecting acquisition curves that can be applied to military tasks. The conditions he identified which impact learning acquisition curves are summarized below.

1. The learning of multi-component tasks (common in military training) may cause temporary asymptotes or "plateaus" in acquisition curves. These temporary plateaus may be due to shifts in the focus on varying components of the task being learned. Over the training period, trainees may shift their attention from easier components (mastered early) to other more complex components (mastered later). These shifts may be reflected in the occurrence of temporary plateaus in skill attainment. The integration of components - as described by Spears (1983) in the process of skill integration - may also play a role in the occurrence of temporary plateaus.

2. Multi-component or complex tasks may cause a great deal of variation in individual acquisition curves.

3. Differences in ability may affect the shape of the acquisition curve. Both asymptote and rate of learning may vary in individuals learning the same task.

4. The degree of prior learning and different training methods may affect the shape of the acquisition curve.

According to Lane (1986b), the study of learning curves is important because it can reveal ultimate skill proficiency, determine when sufficient training or practice has been reached, and track an individual's performance or progress. His conclusions that are relevant to this discussion are summarized below.

1. "The 'typical' curve relating training performance to practice has a characteristic negatively accelerated shape." Curves deviate frequently from that shape for a variety of reasons and due to differing conditions (discussed above).

2. While the general shape of the curve can be reliably anticipated, the "<u>time course</u>" over which acquisition runs - and as a consequence- its parameters cannot be reliably predicted from prior knowledge of task characteristics alone. The mathematical description of a task learning curve requires data specific to a task or training segment.

We share some of the concerns raised by Lane (1986b), particularly those which:

(a) relate to constraints on military training (e.g., the need to train everyone in fixed time intervals tailored more to economics than to the attainment of skilled performance),

(b) confusion in the field regarding the application and meaning of learning curves based on individual data versus group data, and

(c) the masking effects which tend to occur in the context of multi-component tasks.

As a consequence, a HRGM should take these issues into consideration. Constraints on military training, however, appear to have more to do with the judicious use of the growth model than upon any model-intrinsic factors. The question of learning curves based on individual versus group data can be managed within the context of alternatives built into the growth model. For instance, users of the model should be cautioned against the inappropriate application of curves based on group data when only individual data apply. Group curves have their place in the growth model since many projections of interest would be relevant to trade-offs among varying levels of soldier aptitude (e.g., AFQT - Armed Forces Qualification Test - scores and mental category). Finally, the issue of multi-component tasks can only be addressed once the growth model for single-component tasks has been developed. This issue would require additional research to determine the effects of multi-component tasks on the shape of learning curves.

Other Approaches to Predicting Human Performance

In addition to curve fitting techniques, some researchers have attempted to predict human performance using models. Human performance modeling has been used to provide estimates of training time and performance levels with sample data (Schneider, 1990). One innovative approach which applies to modeling was developed by Kennedy and his colleagues (Kennedy, Jones, & Baltzley, 1988a, 1988b). Kennedy et al. developed a model known as the isoperformance model. This model makes assumptions about the relationship between training variables, personnel variables (e.g., aptitude), and system performance and divines an optimum performance level (thus the word 'isoperformance') among these variables.

Isoperformance is based on system "trade-offs"; that is, how a level of operational performance (or minimum cost) can be obtained through different combinations of personnel (i.e., individual differences), training (e.g., number and length of trials), and equipment (e.g., varied features on a single machine or different engineering options). Isoperformance also refers to a curve defining the relationship of training time and individual ability. In other words, an isoperformance curve shows the amount of training time needed by persons of differing ability to become proficient on a system. For example, as seen in Figure 3, persons with higher aptitude require less training time to become 80%, 50%, or 20% proficient on a system than persons with lower aptitude. Although most isoperformance curves are negatively accelerated, they can have other shapes depending upon the distribution of individual aptitudes. For example, in the military, many job specialties require the exclusion of low aptitude personnel. This causes the isoperformance curve to drop sharply at first and then descend in a roughly linear manner (Kennedy et al., 1988b).

According to Kennedy et al. (1988a, 1988b), isoperformance can be used in a variety of ways. For example, isoperformance curves can be used to assess what aptitude category is needed to fill certain positions in the military. Isoperformance can also be used to evaluate current systems (i.e., to suggest improvements or to compare the costs of different parts of the system).

The isoperformance model appears to be an application of learning curves to trade-offs among known system variables (i.e., training, individual differences, and system performance). It appears to add nothing new to the development of valid learning curves themselves. It does appear to serve well in the context of a series of conceptual tools on which to apply empirically-based learning curves, once they are available. Presently, this

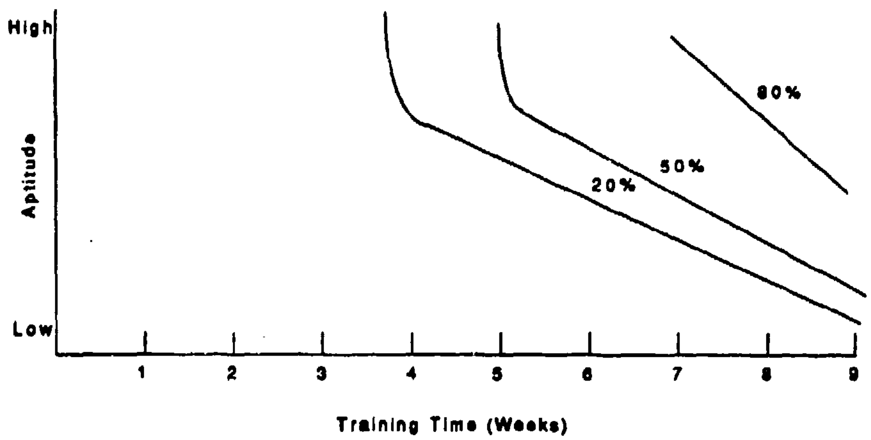


Figure 3. Isoperformance curves for 80, 50, and 20% proficiency (adapted from Kennedy et al., 1988).

line of research suffers from a distinct lack of systematic data.

Issues Relevant to the Measurement and Prediction of the Effects of Training on Performance

A number of related topics arise when discussing the measurement and prediction of the effects of training variables on the performance of military tasks. Some of the more important include:

- Transfer of Training (TOT)
- Training Effectiveness Evaluation
- Performance Measurement
- Experimental and Quasi-experimental Designs
- Simulator Fidelity
- Task and Behavioral Taxonomies

While each of these bear important tangential relations to the topic, it is beyond the scope of this report to describe in detail the importance of each to the measurement and prediction of human reliability growth as defined in this report. Instead, a summary of each topic, including a definition of each topic and its general ramifications to the problem domain, is provided. A topically arranged bibliography of the major research in each area is enclosed as Appendix E.

Transfer of Training (TOT)

TOT, or more generally transfer of learning, occurs when the learning of one task(s) generalizes to other, similar[5] tasks. According to Valverde (1973), "Anything which the trainee can learn can be transferred, including skills, facts, learning sets, self-confidence, interests, and attitudes. Transfer may be either positive or negative. Positive transfer takes place when the acquisition of one skill facilitates the acquisition of another skill. Negative transfer occurs when the acquisition of one skill interferes with the acquisition of another skill."

Derived measures of TOT are numerous, including:

- Percent of transfer (Gagne, Foster, & Crowley, 1948)
- Incremental transfer effectiveness (Roscoe, 1971)
- Transfer effectiveness ratio (Roscoe, 1972)

Other criterion-referenced performance measures often used in TOT studies are listed under the performance measurement topic below.

Lane (1986b) notes that the ultimate criterion of training effectiveness is the capability to perform a job. Transfer is (along with retention of learning) a major intermediate criterion of successful training. The relationship between acquisition of skilled performance and the application (meaning: successful transfer of that skilled performance) are not "straightforward," in Lane's own words. Some training methods have positive effects on transfer simply because they allow the emergence of higher levels of training performance and learning. However, there are trade-offs. Over-training (over-learning) can reduce transfer in some situations. This may be due to the automaticity that occurs with overlearning skills through too much practice. Students may lack the flexibility to see subtle cue or context differences among otherwise similar tasks, that may require differences in their responses. Witness the accident at Three Mile Island (Rogovin & Frampton, 1979), in which control room operators

[5]It is the word, similar, that draws the largest debate among military training researchers. Theories for 'how similar' (meaning how generalizable) exist without general consensus (e.g., Osgood's 1949 transfer surface paradigm), but empirical evidence for similarity (read: generalizability) which supports any theory to date is lacking.

followed their training perfectly - for the wrong accident situation. They failed to properly interpret the cues that would allow them to properly diagnose the true plant status. Correct diagnosis would then have led them to perform the correct procedures for plant management. Other examples exist, but it should suffice that the relationship between TOT and the relationship of training variables and performance is not well defined.

Training Effectiveness Evaluation (TEE)

TEE is a methodology for evaluating the effectiveness of training, primarily training devices and simulators. According to Jeantheau (1971), the following techniques may be applied in the context of a TEE:

- Qualitative Assessment
- Non-comparative Measurement
- Training Device/Conventional Comparison
- Utilization Procedure Comparison
- Training Device Comparison
- TOT

Other researchers also consider cost effectiveness comparisons as a method of TEE.

Only the latter two in the above list, Training Device Comparison and TOT (covered above), are immediately relevant to the question of the relationship between training variables and performance. Usually the training variables evaluated are device characteristics and utilization procedures. Reports which present results of rigorous TEEs may be usable in defining the data base for a HRGM.

Performance Measurement

Performance measurement is the collection of human performance data in either the learning/training environment or on the job. It may be subjective or objective, as well as instrumented (using a computer set up to record performance measures automatically). The relationship of performance measurement to human reliability growth is a potentially negative one. Invalid, inaccurate performance measures can negate the value of the growth model's predictions. It is that simple. Unfortunately, it is not that simple to control the precision of the performance measures.

Lane (1986a) identifies seven criteria for evaluating performance measures that include:

(a) Reliability
(b) Validity

(c) Sensitivity
(d) Completeness
(e) Separability of operator from contributions of the measurement context
(f) Diagnosticity (specificity)
(g) Utility and cost benefit (value against alternatives).

Of these criteria, it appears that a., b., c., and f. would have the most significant implications for the growth model.

<u>Experimental and Quasi-experimental Design</u>

Most TOT studies, TEEs, and applications of performance measurement to the training environment make use of experimental design and statistics. This striving for experimental rigor has also placed a significant expense on the ability to collect adequate training and performance data. While quasi-experimental designs have relieved some of the burden on the experimenter to collect a large enough data set from a large enough sample size, they have done so at the expense of permitting confounds to exist among important training variables. Recent efforts to optimize experimental rigor and to make more economical use of data collection and subject pools include response surface methodology (Simon, 1979) and fitting learning curves using orthogonal polynomials (Sabol, 1986). Sabol's use of orthogonal polynomials is of particular interest because it attempts to make use of the learning curve tendency that occurs in formal training to analyze specific components that can be tested for statistical significance. Unfortunately, the technique does not provide the opportunity for direct determination of the form of the learning curve equation.

<u>Simulator Fidelity</u>

Fidelity in training to the operational environment is a major focus in the investment of training managers for training. Unfortunately, the rules for fidelity levels required are as unknown as the degree of similarity among tasks required for transfer of training to occur. Fidelity is one of the important training variables that a HRGM would hope to predict. Current knowledge in this area does not support a successful outcome with respect to this goal.

<u>Task and Behavioral Taxonomies</u>

The use of task or behavior classification schemes is an intrinsic part of the construction and operation of a HRGM. It represents the primary means by which the learning and performance data and relationships could be accessed for use in making projections. A comprehensive review of existing task and behavior taxonomies was beyond the scope of this project. Certainly, the development of a completely new and unique taxonomy was deemed highly undesirable and was quickly dropped from further consideration, particularly in light of the uncertainties which exist with regard to the ability to collect sufficient quantities of adequate and usable training and performance data on which to develop a growth model. The following three references were reviewed for application to the taxonomic concerns of this project:

O'Brien, Simon, & Swaminathan (1989) Development of PER-SEVAL Performance Shaping Functions.

Hritz (1979) Behavior Taxonomies and Training Equipment Design: A Literature Review and General Model.

Bownas & Cooper (1978) Review of Reported Behavior and Performance Taxonomies.

The taxonomy chosen was contained in the O'Brien et al. report, as will be discussed in a later section of this chapter.

Summary of the Findings Based on the Review of the Theoretical Literature

The review of the literature presented above provided the basis for the following four objectives:

1. Identify the metrics and methodology for use in the development of a HRGM to predict the effect of training variables on soldier performance.

2. Identify the tasks (i.e., taxa) on which to obtain and classify training and performance data required to support Objective #1 above.

3. Identify the kinds of data required to support Objective #1 above and potential sources of such data.

4. Develop the criteria for determining the acceptability of actuarial data for use in constructing a proposed HRGM.

Each of these objectives is discussed in the remainder of this chapter.

Metrics and Methodology for Use in Developing a HRGM

The metrics deemed appropriate for use in the development of a HRGM that predict the relationship of training variables to soldier performance are, not surprisingly, the typical learning curves already reviewed. In general, three of the four major types of learning curve equation forms are considered most useful as metrics to employ in such a model. These are: (a) the power curve, (b) the hyperbolic, and (c) the exponential. The logistic function is the one form of the learning curve equation that may not be appropriate. This is due to our agreement with the literature (Lane, 1986b) that the logistic function may occur only when the source of data is subjective input from subject matter experts (SMEs) such as instructors. The logistic function may map only the SMEs interpretation of student learning acquisition rates. As a consequence, we will not consider the logistic function as a metric to be considered as part of the growth model.

Among the four parameters, identified by Spears (1983) as important aspects of learning curves to consider in evaluating various aspects of learning and performance, two parameters appear to dominate. These include the rate of learning and asymptote parameters. Data which measure these parameters would be most valuable in evaluating the effects of training variables on soldier performance. The other two parameters espoused by Spears (1983), beginning learning and inflection point, are considered less valuable for two reasons. First, the literature does not reflect any known applications of these two parameters. Second, their range of potential application to problems of predicting the effects of training variables on soldier performance is narrow, which makes them of limited value to a growth model.

The types of measures used in the literature for the measurement of both training (independent variables) and performance (dependent) are as shown in Table 1. These measures appear the most likely and more suitable to obtain for use in constructing a HRGM.

The considerations discussed above provided the basis for collecting literature which reports actual findings at a level of resolution usable for constructing a model. Once obtained, the data could then be reviewed and those that fit the project requirement selected. This subset of data would then be analyzed, and a "best fit" learning curve could be developed. A "best-fit" learning curve is one that accounts for the most variance in the data or the highest r^2 value (r is the correlation coefficient). One way to examine the fit of a learning curve to the data would be to transform the dependent and independent variables logarithmically into log-log coordinates using the power function. It is also possible to transform only one of the variables logarithmically into a semi-logarithmic or semi-log coordinates (the exponential function) (Lane, 1986b).

Table 1
Measures Used for Both Training and Performance

A. Measures Used to Assess Training Performance.

- Time to learn
- Rate of learning
- Errors committed
- Performance time on trials/practice
- Asymptote level of learning
- Training test scores (e.g., on job knowledge tests)
- Instructor ratings
- Simulator performance data (instrumented data)
- 'Schoolhouse' grades

B. Measures Used to Assess Job Performance

- Time to perform (and task-based variations)
- Accuracy (and task-based variations)
- Instrumented performance data

The curves derived from the data could then be used in a growth model to form specific trade-off analyses and/or "what-if" type analyses based on the relationship between training and performance depicted by the curves. The curves could also be used to make specific predictions concerning a particular task type (or taxa). For example, if data were gathered from the training of perceptual-motor tasks in addition to cognitive skill tasks and motor tasks, then the perceptual-motor data would not only be useful in making general predictions about training and performance but would also be useful in specific predictions concerning training and performance with perceptual-motor tasks.

Identification of Taxa for Inclusion in the Library of Learning Curves Used in the HRGM

A HRGM requires the use of a series of taxa (meaning globally classified behavioral tasks with strong military relevance) in order to organize and access the learning curve functions which might be available. This would facilitate the use of the growth model in predicting the effect of training variables on performance by providing ease of access to the empirically-based learning curve algorithms contained in the model. While the development of a formal taxonomy of behaviorally anchored taxa was deemed as outside of the scope of this effort, especially since several task and behavioral taxonomies are already in existence (e.g., Berliner, Angell, & Shearer, 1964), there was a need to provide a task-based structure to the collection of relevant learning curve data. Several behavioral taxonomic models were reviewed (O'Brien et. al., 1989; Hritz 1979; Bownas & Cooper, 1978) and compared for utility to this project. It was decided to select the basic taxonomic listing used by O'Brien et al., since this selection would make it feasible to extend the use of the data collected in this project for application to the ongoing ARI research program to develop a family of computer simulation models. These data would have greatest bearing on the PER-SEVAL model member of this family of computer models. The taxa contained in the O'Brien et al. research are shown in Table 2.

Table 2
PER-SEVAL Task Taxonomy

Type	Taxon
Perceptual	Visual Recognition/ Discrimination
Cognitive	Numerical Analysis
	Information Processing/ Problem Solving
Motor	Fine Motor-Discrete
	Fine Motor Continuous
	Gross Motor - Light
	Gross Motor - Heavy
Communication	Oral
	Reading and Writing

O'Brien et al. (1989) state: "The PER-SEVAL task taxonomy is primarily an expansion of Berliner's (1964) task taxonomy." Several of the Berliner' taxa were eliminated from the final list employed by O'Brien et al. on the basis of a lack of military relevance (i.e., tasks which occur either not at all or infrequently in the military context).

The primary application of the PER-SEVAL task taxonomy was to guide the literature search and review activities which made up the predominate effort in this project. It also provided the basis for classifying data. All subsequent descriptions or listings of the data described in this report are in accordance with this basic task taxonomy.

Data Requirements and Sources of Training and Performance Data

In order to apply curve fitting techniques to assess the relation of training and performance, relevant data were required from a variety of sources. These sources could potentially include:

- Military and non-military social and behavioral science databases (e.g., PsycINFO, ERIC, DTIC, NTIS, NASA),
- Data from other sources such as personal libraries and unpublished reports,
- Computerized data bases of training and performance data maintained by the major DOD behavioral and social science laboratories (TPDC, AFHRL, NPRDC, ARI), and
- Collection of new training and performance data for use in developing new learning curve relations.

Of the four types of data sources, this project was limited to the first three. Collection of new data was beyond the scope of work of this project. The search techniques used in obtaining these data are discussed in the next chapter. The issue of data collection to support this line of research is covered in Chapter 4. Chapter 4 also covers conclusions and recommendations.

Criteria for the Acceptability of Data Used in the Development of a HRGM

The criteria to be applied to the data to be considered for use in the development of a HRGM are presented in Table 3. These criteria, derived from our understanding of the underlying research literature on human learning and practice, the use and application of learning curve fitting techniques to the military training context, and the nature and sources of potential training and job performance data, are intended to assist us in answering the basic question:

> Do sufficient quantities of suitable data exist on which to develop a credible HRGM?

An affirmative answer to this question then leads to the conceptualization of the model parameters and constructs which would be incorporated into such a model. A negative answer to this question, however, leads to the conclusion that without such suitable data, such a stand alone model cannot be developed based on the available empirical data. This question and its outcome form the basis for the remainder of this report.

Table 3
Criteria for Evaluating the Adequacy of the Data for Use in Developing an HRGM

Existence of actuarial data, which meet the following subcriteria:
• Measurement of both predictor (measures of training) and the criterion (job performance)
• Measurement that preserves the accounting for individual differences or of meaningful groups (blocks) of subjects (e.g., AFQT category)
• Inter-study comparability of independent and dependent variables within taxa
• Same independent and dependent measures (or transformable to same independent and dependent measures) employed across studies
• Sample size of study from which data are derived
Relevance to military task requirements
Relevance to military training environment
Comprehensiveness of Taxa coverage

Methodology for the Collection and Reduction of Quantitative Human Performance Data Relevant to Task Training

The collection and reduction of quantitative human performance data was required in order to answer the question previously posed in the last chapter, that is;

> Do sufficient quantities of suitable data exist on which to develop a credible HRGM?

The initial effort to answer this question involved a meta-analytic model. As will be explained, this approach had to be abandoned and a less rigorous approach was devised in which suitable data were collected in whatever form they were presented and classified within the PER-SEVAL taxonomy (O'Brien et al., 1989). The discussion that follows presents both of these approaches as well as a detailed description of the literature review and results obtained.

Initial Approach: Meta-Analysis

The first step in developing a HRGM was to derive quantitative relationships of the impact of training on performance from the human behavioral and social science literature using meta-analysis. Meta-analysis is a statistical procedure for combining research results of a 'similar nature and kind' across numerous studies.[1] For example, meta-analysis has

---

[1] Meta-analysis begins with the formulation of questions or hypotheses concerning some variable of interest. In our case, the variables of interest would be training variables, such as:

- different methods of instruction,
- different equipment, part task versus whole task training devices, and
- training time.

The key to this effort is the effect of these training variables on human performance. For example, one approach would be to study the effect of part task versus whole task training on the learning of a visual discrimination/information processing task such as pursuit tracking. Next, articles relevant to the research question would be collected from various sources including literature searches, journals, and private libraries, and screened for inclusion into the meta-analysis. The screening process would ascertain: (1) whether the articles were relevant to the research question, and (2) whether the articles contained the statistical information needed to conduct the meta-analysis.

The results from each of the articles have to be converted into a common statistic under the rules of meta-analysis. As a result, each article should report the results of statistical tests such as the descriptive statistics of $\underline{t}$, $\underline{F}$, x^2, r^2, or, at a minimum, the mean, standard deviation, and sample size. The next step would be to convert the results from each article into a common descriptive statistic. One of the most widely used descriptive statistics in meta-analysis is the index of effect size or the "d" statistic (Cohen, 1977). The d statistic is a transform of the point-biserial correlation (Hunter & Schmidt, 1990). Specifically, it is the difference between the means of the control and experimental groups divided by the within group variance (i.e., the sum of the variances of the experimental and control groups divided by 2). However, other statistics may be used in meta-analysis such as the point-biserial correlation r and the been successfully applied to determine the effectiveness of psychotherapy (Glass, 1977) and to examine the relationship between spatial experience and spatial test performance (Baenninger & Newcombe, 1989). The results of each meta-analysis would be organized by a predefined set of globally classified tasks or taxa with a strong military relevance (i.e., the PER-SEVAL taxonomy by O'Brien et al., 1989).

However, our initial approach soon wavered and was subsequently revised because several problems emerged which affected the feasibility of using meta-analysis to develop the model. A major problem was the tremendous diversity of task types (e.g., some tasks were composed of single component operations while others were multi-component) found in the data within each of the taxa. In addition, the types of training used within and between the taxa were as diverse as the task types were within. Both of these problems made logical groupings of related experiments virtually impossible to derive. Another problem was the uneven distribution of studies for the taxa; there were no studies representing the taxa of numerical analysis, gross motor-light, and communication-oral while the taxa of visual information processing and information processing were relatively data-rich. There was also a problem in how to incorporate multiple independent variables within one research article into the meta-analysis. Thus, it was concluded that a meta-analysis of the impact of training variables on performance organized by behavioral taxa could not be supported by the extant data.

Our problems in obtaining suitable data to support the meta-analysis are not altogether surprising because similar problems have been reported elsewhere by Kennedy et al. (1988a; 1988b). Specifically, Kennedy et al. surveyed the systems research and human factors engineering literature for studies comparing equipment, aptitude, and training. Of approximately 10,000 articles searched, only 276 involved studies of training and

---

the variances of the experimental and control groups divided by 2). However, other statistics may be used in meta-analysis such as the point-biserial correlation r and the population strength between the independent and dependent variables or omega squared (Hays, 1973).

The next step in the meta-analysis would be to test the effect sizes of the studies for homogeneity of variance. If the variances were homogeneous, then a combined effect size would be computed. The combined effect size is a sum of the effect sizes divided by the number of studies. Finally, the null hypothesis would be accepted or rejected depending on whether or not it fell between prescribed confidence intervals (for a more detailed description of meta-analysis, see Glass, 1977; Hunter & Schmidt, 1990).

The results from the meta-analysis would then be used to develop a model or algorithms to predict the impact of training variables on performance. Predicting the effect of training variables on performance would be accomplished by either employing curve fitting techniques to derive taxa-based learning curves or by building regression models.

performance as a function of equipment variations. From this, only 10 articles reported information necessary to calculate the combined inferential statistic, omega-squared. Although there were no problems in calculating omega-squared in each study, Kennedy et al. (1988b) reported that the data "turned out to be too irregular to permit sufficient generalizations about trends in these studies" (p. 59). In a similar vein, Lane (1986b) reports that published research on acquisition and learning could not support generalized predictions concerning the time course of training across different training situations with learning curves.

Revised Approach: Generalized Classification of Learning and Performance Literature

Our revised approach, therefore, involved locating suitable research articles which presented data and or learning curves presenting the relation of training to performance. Essentially, the technical approach involved collecting and reviewing data from the literature and, if feasible, analyzing the data to derive learning curves. The curves derived from the data would then be usable in a growth model to perform specific "trade-off" and "what-if" analyses based on the relationship between training and performance depicted in the learning curves. No special assumptions were made about the data obtained at this point in order to permit the widest possible 'net' in our data capturing effort. The discussion that follows presents the approach used and results obtained during the literature search.

Initial Literature Search

The difficulty in identifying suitable sources of human performance data related to training is underscored by the difficulties encountered in performing the literature search. Two efforts were made to use automated literature search techniques. The initial literature search was structured to obtain a list of publications relevant to TRAINING *and* HUMAN PERFORMANCE in accordance with each of the taxa contained in the PER-SEVAL taxonomy. Table 4 summarizes this search strategy.

The literature search commenced with a review of the following on-line databases using this literature search strategy:

- Defense Technical Information Center (DTIC),
- National Technical Information Service (NTIS),
- Educational Resources Information Center (ERIC),
- Psychological Abstracts Information Services (PsycINFO), and
- SOCIAL SCISEARCH.

Table 4
Keywords Used in the Initial Literature Search

"Training" and "Performance" and each of the following individually:
Visual Recognition/Discrimination
Numerical Analysis
Information Processing
Fine Motor (discrete)
Fine Motor (continuous)
Gross Motor (heavy)
Gross Motor (light)
Communication (read and write)
Communication (oral)

Not only were the selected taxa used as keywords, but they also became the principal means of classifying the relevant literature identified through the literature search.

A print-out of titles (over 3,000 separate titles in our estimate)[2] of literature citations derived from the search was obtained and approximately 500 citations were identified as being candidate publications. Among the citations were articles from the broad behavioral science and human factors literature relevant to training and human performance and at least one of the nine taxa listed above. A further request was made to obtain a print out of the abstracts of the 500 candidate articles. These abstracts were then carefully evaluated in terms of:

- apparent probability of the presence of quantitative data,
- appropriate statistical design and analysis,
- proper population (i.e., normal adult population), and
- (to a lesser extent) content.

This process resulted in the selection of 108 acceptable abstracts.

The next step involved the collection of publications. Attempts were made to obtain the 108 candidate publications from

[2] It was difficult to determine the number of original "hits" in the literature, since there were so many repetitions of titles across search terms and data bases.

several sources, including the ARI Headquarters Technical Library and several local university libraries. The candidate publications were then evaluated in terms of appropriateness of the data, statistics, and content. This procedure resulted in the selection of 42 articles. Moreover, this stage marked the establishment of the visual and information processing taxa as the most data-rich in terms of available relevant literature (i.e., all but 6 of the initial 42 studies examined were classified as either visual recognition/discrimination or information processing).

Expansion of the Literature Base

These initial 42 articles were summarized and presented in an In-Process Review to ARI scientists. Upon further examination, it was concluded that only 5 of the 42 articles contained predictor and criterion variables appropriate to the project. After discussion with ARI scientists, it was decided that an expansion of the literature base was a necessary next step in order to assure ourselves that the search strategy was not the cause of the failure to find larger amounts of suitable data. Two initiatives commence to expand the literature base:

- search of other government laboratory data bases, and
- development and implementation of new keywords in future searches.

Search of government data bases. Major government laboratories were contacted to determine the existence and availability of relevant data, including:

- Training and Performance Data Center (TPDC),
- Naval Personnel Research and Development Center (NPRDC),
- Air Force Human Resources Laboratory (AFHRL), and
- Army Research Institute/Manpower and Personnel Research Laboratory (ARI/MPRL).

Summaries of the discussions with each organization are presented below.

Points-of contact at TPDC expressed doubt as to the existence of training and performance data relevant to the project. It was suggested, however, that we contact ARI/MRPL representatives for possible data. Subsequent discussion with MPRL personnel informed us that only attitudinal data were presently being collected at ARI (which would not be relevant to this project).

Representatives of NPRDC were also contacted with respect to the availability of training and performance data. It was revealed that NPRDC had an extensive training data base, but job performance data were not collected. Additional conversation with NPRDC representatives in San Diego disclosed that training and performance data had been collected, but several methodological confounds precluded the usefulness of the data for our purposes.

Continued efforts were made to obtain information pertaining to the availability of training and performance data from government laboratories. Contact with AFHRL in San Antonio revealed that the Air Force has ongoing training projects with emphasis on planned instruction, but no appropriate data have been collected.

Implementation of new keywords. The second method by which attempts were made to expand the literature base involved the implementation of new keywords derived from the 5 articles containing predictor and criterion variables. The new keywords were used to peruse areas of the literature that were unexplored in the initial search. These additional keywords are shown in Table 5. As a preliminary examination, these keywords were used in an interactive search of the on-line data bases of ERIC and PsycLIT (a compact version of PsycINFO). This procedure was used to determine the "hit-rate" of each keyword so that this information could be used as a guide in preparing for the primary search.

Table 5
Keywords Used in Subsequent Literature Search

Training
Learning
Performance
Asymptote
Training and Learning Performance and Asymptote
Skill
Transfer
Learning and Skill Transfer
Transfer of Training
Military
Transfer of Training and Military
Skills
Skills Training and Military
Military and Training

Through conversations with AFHRL, it was suggested that we consider conducting future literature searches via Crew System Ergonomics Information Analysis Center (CSERIAC; a service organization that can search large government laboratory databases). Ultimately, a request was made of CSERIAC to conduct an additional search of the DTIC database as well as a search of the NASA database using the new keywords. All publications obtained as a result of this search were reviewed and 14 articles were determined to be relevant. Thus, at this point we had obtained a total of 19 relevant articles (14 new articles combined with the 5 obtained in the initial searches).

Non-traditional search techniques. A final attempt at expanding the literature base was initiated using non-traditional search strategies (also known as "brute force techniques"). In effect, the search involved reviewing relevant articles from personal sources (e.g., current journals including Human Factors, selected proceedings of the Human Factors Society, reference sections of books, reference sections of articles). In addition, Volumes I and II of the Handbook of Perception and Human Performance by Boff, Kaufman, and Thomas (1986) were examined. Following this search, 14 prospective articles were obtained from local university libraries and these articles were evaluated. After evaluating the articles, 8 were determined to be relevant to the project. Thus, ultimately we were able to obtain a total of 27 relevant articles (5 from the initial on-line search, 14 from the subsequent on-line search, and 8 from personal sources; see Table 6 below for pertinent characteristics of the relevant articles).

As seen in Table 6, the Information Processing taxonomy and the Visual taxonomy were the only taxons in which articles were identified as being data-rich and data-relevant. It may be that the areas of Information Processing and Visual Performance are research areas in which there is more emphasis on the interaction of training and human performance. If this were the case, then it would partly explain our ability to obtain more literature in these areas. However, it is also possible that more non-traditional search techniques might reveal additional data in the Information Processing and Visual Performance areas as well as the other taxa. For a complete list of the distribution of relevant articles per taxa, see Table 7.

Table 6
Characteristics of Relevant Articles

TAXA	#	Author & Date	Independent Variables	Dependent Measures
INFO. PROC.	1	Basadur et al. (1982)	Experiential or traditional training in creative problem solving	a) Self-report attitudinal/ideation questionnaire b) Product planning notes of subjects recorded on tape c) Self-reported interviews of subjects observations of job perf. in industrial engineering
	2	Boreham (1985)	Hypothesis testing training for simulated diagnostic tasks used in salt packaging plant	Number of different categories of hypotheses generated during pretest and posttest
	3	Briggs et al. (1965)	a) Interactive vs isolated trng of team members in radar control b) Complexity of radar control task c) Length of training for radar control	Efficiency score composed of the number of successful intercepts, time, and fuel consumed
	4	Card et al. (1978)	a) Type of Device (mouse, joystick, step keys, text keys) b) target size (# of letters of text) c) Distance from starting point on all positioning devices	a) Positioning speed b) Homing time

Table 6
Characteristics of Relevant Articles (Continued)

TAXA	#	Author & Date	Independent Variables	Dependent Measures
INFO. PROC. Cont.	5	Fisk et al. (1988)	a)Memory set size b)Possibility of correct decision (positive or negative trials)	a)Reaction time in a dual-task situation
	6	Johnson (1978)	Initial training strategy for "checklist" sequential task: conventional, reproduction, or blind strategy	a)Sequence and setting errors on sequential task b)Time to perform last trial of sequential task c)Transfer errors on actual sequential task (e.g., control panel equipment)
	7	Kanfer et al. (1989)	a)ATC part-task trng strategy: declarative or procedural b)Goal setting strategy group: goal vs no goal instructions c)Ability via ASVAB score	Number of planes landed on six trials of full ATC
	8	Malloy et al. (1978)	Cognitive trng strategy group: using Raven's Progressive Matrices Test, using brief exposure to test, or no exposure	a)Problem solving performance on 17 matrix puzzles b)Piagetian multiplicative classification test
	9	Masson (1986)	Degree of exposure to word triplets	Word identification score

Table 6
Characteristics of Relevant Articles (Continued)

TAXA	#	Author & Date	Independent Variables	Dependent Measures
INFO. PROC. Cont.	10	Miller (1975)	Instructional strategy for Improved Hawk: traditional vs supplemental	a)Practical quiz test score on Improved Hawk b)Standardized score on Improved Hawk
	11	Morris et al. (1985)	Instructional training for dynamic production process using simulator (PLANT)	Production measured on dynamic production process simulator (PLANT)
	12	Myers et al. (1987)	Mapping of info. (variably or consistently mapped) in trng for telecomm. visual search task	a)Reaction time on telecomm. visual search task b)Accuracy on telecomm. visual search task
	13	Patrick et al. (1988)	Trng materials: technical story vs heuristics, in fault finding tasks	Accuracy and speed in fault diagnosis test in transfer scenario
	14	**Roberts et al. (1983)	Experience with text editors: novice vs expert	a)Time-score for each text editor b)Individual error score for each text editor
	15	Rouse (1978)	a)Problem size in fault diagnosis using graphically displayed network system as simulation b)Pacing: forced vs non-forced c)Display: computer aided vs non-aided	a)Number of tests before correct diagnosis b)Percent correct

Table 6
Characteristics of Relevant Articles (Continued)

TAXA	#	Author & Date	Independent Variables	Dependent Measures
INFO. PROC. Cont.	16	Scott et al. (1982)	Use of SH-3 Aircraft Training aids: traditional or supplemental	Errors on cockpit procedures trainer (CPT)
	17	Sheppard (1984)	a) Task configuration: part vs whole task b) FLOLS type: conventional or vertical bars added c) FLOLS size	RMS Glideslope error
	18	Weitzman et al. (1979)	Type of Trng instrument: Device 2B24 simulator, UH-1H helicopter, or both	Instructor pilot ratings of performance
VISUAL	1	Cockrell (1979)	Target quality during trng: degraded vs normal	Mean number of correct identifications of degraded targets
	2	Eberts (1987)	a) Task type: single or dual b) Augmentation type (point or parabola cues)	RMS tracking perf.
	3	Edward et al. (1979)	Pretraining: multimedia vs no pretraining	Instructor-rated transfer perf. in the T-37 aircraft
	4	Lintern et al. (1990)	Landing instruction: supplemental simulator trng or traditional trng	Raw tracking error on landing task

Table 6
Characteristics of Relevant Articles (Continued)

TAXA	#	Author & Date	Independent Variables	Dependent Measures
VISUAL Cont.	5	*Rabbit et al. (1979)	Length of practice on visual search of embedded letters	Reaction time for visual search of target letters
	6	Schneider et al. (1984)	a)Exp. 1: Type of mapping: variably vs consistently during trng for category search task b)Exp. 2: Number of concurrent tasks: Single or dual	a)Exp. 1: Visual search reaction time improvement b)Exp. 2: Percent improvement of word detection
	7	Simon et al. (1981)	Simulator configurations (variations of: control order, display lag, tracking mode, prediction time, control gain, and number of trials)	Tracking error as measured on the Visual Technology Research Simulator
	8	Wightman et al. (1987)	a)Trng task: whole vs segmented task b)Carrier landing lag: normal lag vs progressive lag	RMS carrier landing as measured on the Visual Technology Research Simulator
	9	Wrisberg et al. (1983)	Trng condition: open skill (requiring varying reactions) vs closed skill (requiring stable reactions)	Absolute error of arm movement during transfer

NOTE: Unless marked as shown below, all studies contain only group data
* Study contains only individual data
** Study contains both individual and group data

Table 7
Distribution of Relevant Articles Obtained Per Taxa

DISTRIBUTION OF RELEVANT ARTICLES OBTAINED PER TAXA	
Information Processing	**18**
Numerical	0
Visual	**9**
Fine Motor (Discrete)	0
Fine Motor (Continuous)	0
Gross Motor (Heavy)	0
Gross Motor (Light)	0
Communication (Read/Write)	0
Communication (Oral)	0
TOTAL	**27**

A textual and graphical summary of each of the 27 relevant articles can be found in Appendix A. A list of citations including only these 27 relevant articles can be found in Appendix B. In effect, Appendix B is intended as a bibliography of articles contained in Appendix A. For an annotated listing of articles obtained and reviewed during all the searches, see Appendix C. Appendix C, thus, is intended as a bibliography of all publications that were accessed through the traditional keyword as well as the non-traditional searches.

Over the course of searching the literature for publications specifically examining the quantitative relationships between training and soldier performance, it became apparent that the traditional keyword search technique was ineffective. The keyword approach allows one to easily obtain traditional literature pertaining to learning as well as traditional literature pertaining to human performance and training. However, it appears that this approach does not effectively lend itself to directly accessing literature common to all of these areas. This is evidenced in the fact that more publications were obtained via non-traditional search strategies than in the initial keyword search.

Feasibility of Developing a Stand Alone Human Reliability Growth Model

The overall feasibility of developing a stand alone HRGM depends upon satisfying the following issues:

1. Do sufficient quantities of suitable data exist?

2. Does an adequate theoretical basis exist for developing such a model, data adequacy aside?

3. Is a stand alone quantitative model feasible, or should the resultant algorithms be incorporated within the structure of a human performance simulation model?

This chapter addresses each issue in turn. The first section of this chapter addresses issue #1. The criteria for the adequacy of actuarial data presented at the conclusion of Chapter 2 are applied to this question. The middle section of this chapter addresses issue #2. The body of knowledge regarding human learning and practice and the learning curve construct is revisited to address the nature of a HRGM. The final section of this chapter addresses issue #3. Given the resolution of the first two issues, recommendations are addressed for the development of a stand alone quantitative model versus incorporation within ongoing efforts to develop human performance simulation models. An illustration of a 'successful' algorithm is also presented in order to describe the ideal situation.

Do Sufficient Quantities of Suitable Data Exist?

The issue which predominates throughout this project is the availability of adequate data. Absent the ability to collect new data depicting the relationship of training variables to human performance, we must depend upon the existence of acceptable actuarial data in the reported behavioral and social science literature to provide these data. The criteria presented in Table 3 describe the basis for determining the acceptability of the available data. Table 8 summarizes our findings with respect to our analysis of the data base assembled through the literature reviews conducted during this project (as described in the previous chapter). The paragraphs that follow discuss each criterion in turn.

Table 8
Summary of Findings Regarding the Adequacy of Actuarial Data for the Development of a Stand Alone HRGM

Criteria for Evaluating the Adequacy of Actuarial Data	**Summary of Findings with respect to Analysis of Data Contained in Extant Literature**
1. Existence of actuarial data	• 27 research articles identified which contain data on variables of interest
a. Measurement of Predictor & Criterion	• 27 research articles containing clearly identified criterion & predictor relationships
b. Preservation of individual differences or of meaningful groups of subjects	• 26 research articles presenting group data[1] • 1 research articles presenting individual data
c. Comparability of independent & dependent variables within a taxa	• Wide variation in types of both independent measures & dependent measures within available taxa • Pooling of data within taxa not feasible
d. Same (or transformable to same) independent & dependent measures across studies	• **Independent measures**: 0% comparable 100% not comparable • **Dependent measures**: 26% comparable 74% not comparable

[1] Note: One article contained both individual and group data.

Criteria for Evaluating the Adequacy of Actuarial Data	**Summary of Findings with respect to Analysis of Data Contained in Extant Literature**
e. Sufficient sample size	• Sample size breakdown: **N < 10**: 26% of research articles **N > 10 and < 50**: 48% of research articles **N > 50**: 26% of research articles • Small sample sizes prevalent which undermines stability of the learning/performance relationships contained in the literature • Insufficient stability in learning curves for projected population (Army soldiers)
2. Relevance to military tasks	• 15% relevant to Army tasks • 26% relevant to Military tasks (other than Army) • 59% not relevant to military tasks
3. Relevance to military training	• 15% relevant to Army training • 26% relevant to Military training • 59% not relevant to military training
4. Comprehensiveness of Taxa Coverage	• **Within**: Indeterminate, but believed inadequate for the 2 taxa covered • **Across**: Coverage of 2 of 9 broad taxa contained in the PER-SEVAL model

Existence of Actuarial Data

A total of 27 research articles was selected out of a list of approximately 3,000 articles identified. These 27 present learning data of the type required for this project. Eighteen of the research articles were identified as belonging to the information processing/problem solving taxon. Nine of the research articles were identified as belonging to the visual recognition/discrimination taxon. No research articles whatever were identified for the remaining seven taxa. This compilation of literature represents a broad range of independent and dependent variables. As a consequence, no obvious classification structure was evident as a basis for pooling studies within a taxa. In sum, while data exist, paucity and diversity in the range of variables encompassed overwhelm potential usefulness for the development of a stand alone, quantitative model. The following summarizes the results obtained in terms of five subcriteria.

(1) Measurement of Predictor and Criterion. In general, the research contained in Appendix A should be classified as 'near misses', that is, data are reported within each article which may be used to approximate both criterion and predictor data. Predictor data (i.e., training data) uniformly exist for the literature reported. However, criterion data (i.e., job performance data) must be approximated, because they are not generally presented as such. Proficiency (the ability to perform a task that has been trained through a 'job sample' or simulated performance) is provided as the criterion. Job performance data require the availability of 'external' measures of criterion performance (meaning external to the training environment). As a consequence, the utility of the data must be weighed against the modeler's acceptance of 'proficiency' data as 'job performance' criterion measures.

(2) Preservation of individual difference or of meaningful groups of subjects. As Lane (1986b) maintains, there are differences between the use and interpretation of learning curves based on individual versus group data. Only two of the research articles contained in Appendix B contain individual learning data. Twenty-six of the articles present group data.[2] One article reports only group data. Individual data are preferred since these provide the best basis for determining the variability attendant to task learning within a group.

---

[2] NOTE: One article reports both individual and group data.

Given that few such data were obtained in the literature search, we are dependent upon the use of group data. We have left the question of 'meaningfulness' to the user of the data, since this is a subjective metric dependent upon the context of use and the interpretations of 'meaningfulness' by the analyst. The only criterion used in selecting the research articles to include was that the subject population represent a 'normal adult population' of subjects. Clearly, if 'meaningfulness' involves separate treatment groups distinguished by aptitude then the data reported here do not apply. However, if 'normal adult population' is considered representative of the 'military population' then dependent upon the situation, the data may apply. This question is for the analyst using the data to decide.

(3) Comparability of independent and dependent variables within a taxa. In order to pool data[3] of a single taxa reported in this literature, the data must have comparable independent and dependent variables. That is, they must contain and vary the same training and job performance variables. As shown in Table 6, the literature contains a broad range of independent variables, including:

- type of training,
- complexity of training,
- training materials,
- learning strategy, and
- extent or amount of training/practice.

These disparate training variables lack a common basis for classification in order to permit ready pooling of data. Likewise, Table 6 contains a broad range of dependent variables, including:

- Problem solving,
- Transfer of skill,
- Spatial rule learning,
- Problem solving quiz,
- Learning score,
- Simulator performance, and
- Tracking and visual discrimination performance.

The dependent variables also lack a common basis for pooling. No two studies have both comparable

---

[3] Pooling of data appears necessary in order to obtain the most robust and comprehensive basis for developing and applying learning curves for use in a model.

independent and dependent variables on which to pool data.

(4) Same (or transformable to same) independent and dependent measures across studies. Even in cases where the variables are the same (or similar) the measures used to operationalize the variables must also be the same (or similar) in order to pool the data across studies within a taxa. Similar training or performance variables identified in the literature, lack similar operational measures. For example, in the case of independent variables, the number of training trials is not readily transformable to training time. (Such a conversion would require knowledge of time elapsed per trial.) In the case of dependent measures, operationalized measures of time or accuracy are dissimilar from one another. Lacking a common yardstick of comparison results in an inability to pool data within a taxa.

(5) Sufficient sample size. Two training variables (known as Performance Shaping Functions or PSFs) were included in the PER-SEVAL report (O'Brien et al., 1989). These variables are frequency and recency of training. The sample size for each of these variables varied from Taxa to Taxa and tasks within taxa, but the range was from 346 to a high of 694 soldiers per task. Most were above 600 soldiers as the sample size. Only two of the studies described in Appendix A reported samples anywhere near this magnitude. One study reported a sample size of 552 subjects and a second reported 288 subjects. Studies with these sample sizes are in the clear minority. The range in sample size is shown in Table 9. As can be observed from the table, 74% of the articles reported subjects with no better than moderate sample sizes. This would tend to undermine the generalizability of the data to other subject populations. In sum, the reported literature does not generally approach the sample sizes used in similar research efforts.

Table 9
Range of Sample Sizes Across Selected Research Articles

RANGE in # of Subject	Number (and Percentage) in Range
Low (< 10 subjects)	7 articles (26%)
Medium (26 - 53 subjects)	13 articles (48%)
High (54 - 552)	7 articles (26%)

Relevance to Military Tasks

Relevance to military tasks is an important issue (as noted by Lane, 1986b). Research on tasks which are not directly relevant to military tasks and environments may not be generalizable to military situations. Some tasks performed in a non-military setting may have no military counterpart. Of the 27 research articles contained in Appendix A, less than 50% represent military-based research. A simple majority of the research comes from other than military environments. While this does not immediately result in an inability to generalize these data, it does raise the concern of generalizability. Relevance is determined by subject matter expert judgement and is not easily assessed in any other manner. Even if given researchers are willing to accept research findings obtained out of the military context, other users of a HRGM may question the model's validity on this basis.

Relevance to Military Training

In like manner to the above criterion, data on specific training variables not obtained in a military training environment may not adequately generalize to the military situation for two reasons:

- the 'treatments' given to training may not be feasible in the military training context (e.g., training time in the military is more a question of budget and tradition - that is, how much training was provided in the past for this or a similar task; in other non-military training situations these constraints may not exist), and

- the training variable may have no practical alternative in the military context.

The percentages of articles relevant to military tasks apply equally well to military training relevance. In sum, relevance to military tasks and training must both be judged and the effects of acceptance by model users must be considered in these judgements.

Comprehensiveness of Taxa Coverage

The degree to which a given set of data or learning curve relationships depict the 'true' relationship of training variables to performance within a given taxa, is in part a question of the degree to which the data are representative and comprehensive of the relationships theoretically possible in each taxa. For this project, this is indeterminate, because we simply *don't know* what we don't know. This indeterminacy creates problems particularly for sets of learning curve relationships that are based on small or only moderate sample sizes. The concern with the underlying generalizability of the data to a new context is proportionally increased. We not only do not know how generalizable the data are, we also don't know how much error we may introduce as a consequence of using the data.

Across the nine taxa in which sources of data were sought, only two of the nine taxa yielded any reported findings. Clearly, the comprehensiveness of coverage of the range of taxa found in the military settings is partial at best. Any stand alone model which could be developed should contain at least some data for each of the relevant taxa. This clearly was not the case here.

Probable Reasons for a Lack of Suitable Data Reported in the Behavioral and Social Science Literature

The lack of suitable data which would permit the development of a quantitative stand alone HRGM was disappointing. During the course of the literature reviews, reports containing similar objectives to our own were reviewed. Their findings were much the same as our own. That is, the theoretical basis for developing and employing learning curve fitting techniques for evaluating the relationship of training variables to human performance is adequate. The data base supporting such an effort does not exist at this time (see Lane, 1986b and Kennedy et al., 1988a; 1988b for a detailed report of these findings). Table 10 presents a summary of the probable explanations for the lack of suitable data reported in the behavioral and social science literature. We can offer no speculation as to the relative frequency of occurrence of any of these probable reasons for the failure of the literature to support our data requirements. This failure may simply reflect the difficulties and expense involved in collecting data of the type required by this project.

Table 10
Probable Reasons for Lack of Suitable Data

• Empirical studies of human learning do not typically involve predictions to an external criterion (i.e., job performance)
• Initial learning or training apparently has not had wide application as a predictor in industrial/organizational psychology
• Development of good criterion measures of job performance are difficult to secure for a variety of practical or theoretical reasons and intermediate measures of proficiency (results of post-training job knowledge tests) are more convenient to acquire as a substitute
• Research that produces no significant findings tends not to be reported in the literature
• Lack of opportunity or support resources to collect training and more particularly performance data (e.g., no opportunity to manipulate training variables such as an invariant training schedule; lack of control over trials and practice periods)
• Lack of concern among researchers

Does an Adequate Theoretical Basis Exist for Developing a Model?

The 'ubiquitous' learning curve appears to provide a solid technical basis for the development of a HRGM. The summary of documented research on the learning curve provided in Chapter 2 demonstrates the veracity of the learning curve as a driving construct in the development of a HRGM. What remains is to describe the manner in which the learning curve might best be employed within a stand alone model. It is not appropriate to state the exact form that such a model should take with any finality. This statement is based on the following observations from our review of the literature:

1. First and foremost, learning curves are data driven. That is, the exact form of the learning curve function is derived directly from the data themselves. Until such time that sufficiently robust data are produced, the search for the truly generalizable learning curve applicable to specific behavioral taxonomic elements appears to be a search for the 'holy grail'. A broadly generalizable learning curve function appears to be the central requirement of such a model of human

reliability growth.

2. There is no consensus on the 'optimum' form of the learning curve function (i.e., power, hyperbolic, exponential, logistic). Depending on the situation, each function has been demonstrated as 'the best fit' equation.[4]

3. The four specific parameters of the learning curve identified by Spears (1983); beginning level learning, asymptote, rate of learning, and inflection point (discussed in Chapter 2) are of value to the prediction of differing training variables - depending upon the circumstances. For example, asymptote is most useful in determining the optimum training time or practice trials for the learning of tasks. Rate of learning is also useful in this determination, but its focus is towards the determination of 'what happens to task proficiency if I stop trials here' rather than the determination of optimum performance levels based on the achievement of asymptote levels of proficiency. The other parameters have value in model applications as identified in Chapter 2.

4. The nature of task performance in the complex settings of military systems may overly complicate our ability to predict the relationship of training variables on job performance. Tasks which contain multi-component operations may 'be more than the sum of their parts' when applying current learning curve technology. This research is still largely reductionistic and we lack the capability to integrate the effects of numerous training variables (e.g., time, media) on the performance of tasks containing multi-component operations.

Recognizing these limitations, we can nevertheless illustrate the desired characteristics and operation of an 'idealized' algorithm to which the HRGM should aspire. What follows is a description of an 'idealized' example, that is, the 'heat effects' algorithms used by Siegel, Wolf, and Bartter (1982).

The 'heat effects' algorithms developed by Siegel et al. (1982) were used in adjusting the impact of heat on human abilities within the Maintenance Personnel Performance Simulation (MAPPS) developed by the same authors. In nuclear power plant

[4] This may in fact constitute an academic discussion, since there is no inherent reason that the HRGM must be based on a 'single' best fit equation. It simply means that we cannot lay out the 'idealized' HRGM with any degree of finality at this time.

operations (as in other environments), heat has significant impacts upon the performance (both physical and cognitive) of humans. High heat conditions occur in situations when the external temperature increases beyond the comfort and/or tolerance ranges for humans performing 'perishable'[5] tasks. Further, it is known that heat stress causes incrementally greater decrements in task performance requiring mental or cognitive operations than for physical operations (Siegel et al., 1982). We should anticipate, then, that cognitive task performance of humans will experience high levels of stress, reduced capability to perform under conditions of high workload, greater response time, and reduced overall accuracy in task performance when heat stress is a factor. Any model which purports to be inclusive of the major variables that affect human performance in which heat is expected to play a significant role should be expected, therefore, to be sensitive to differences in task performance that occur as a function of heat stress.

Siegel et al. (1982) derived mathematical relationships from the isodecrement functions presented in Figure 4 (A & B) in order to model the effects of heat stress on maintenance task performance in nuclear power plants. They applied the results of two literature reviews (Grether, 1973; Ramsey & Morrissey, 1978) that consolidated the research findings related to the effects of heat on human performance. These studies summarized the findings of 16 cognitive studies, 53 reaction time and cognitive studies, 7 vigilance studies, 14 tracking studies, and 48 perceptual-motor studies in which heat was an independent variable. The integration of the results of these studies resulted in the isodecrement curves contained in the figures. Figure 4 (A) plots the effect of temperature over time on task performance for mental reaction time. Figure 4 (B) plots the effect of temperature over time on task performance combining tracking, vigilance, and complex tasks. From these plots, two regression equations were derived, one for intellective tasks and another for perceptual-motor tasks. These equations are shown in Table 11. These equations were then used to develop ability decrement functions for maintainer performance as a function of temperature over time, as shown in Table 12. The functions were then incorporated within the ability decrement subroutines programmed for the MAPPS model to calculate ability decrements whenever heat is included as an input variable in the simulation.

---

[5] Perishable in the sense that task performance degrades as a direct consequence of heat stress effects on humans.

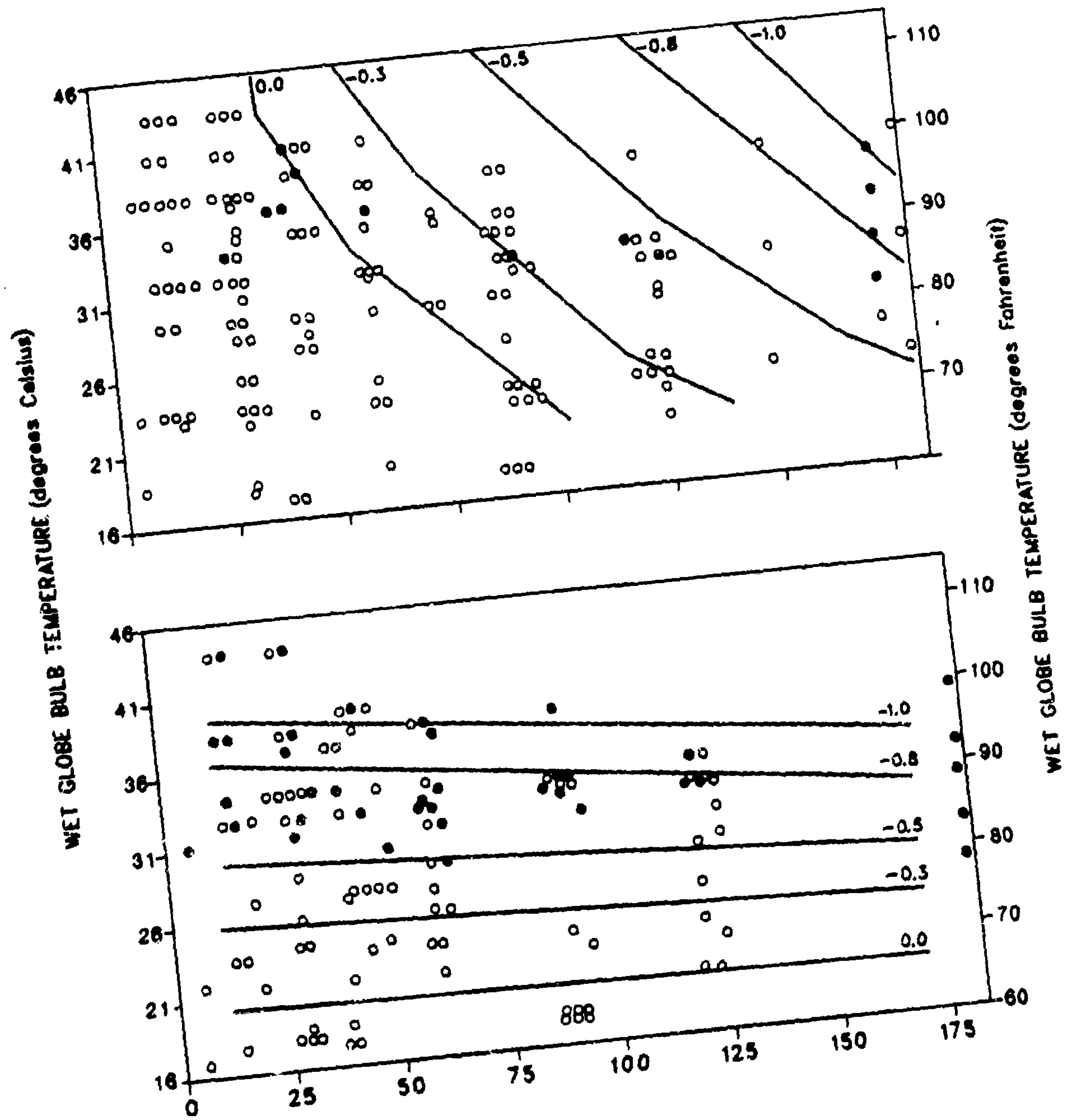


Figure 4 (A & B). Isodecrement functions of Grether (1973) and Ramsey and Morrissey (1978).

Table 11
Regression Equations Based on Isodecrement Functions

INTELLECTIVE TASKS:

$$Y = .923 + .002(\text{TIME}) - .0054(\text{TEMP}) - .0008(\text{TIME} \times \text{TEMP})$$

PERCEPTUAL-MOTOR TASKS:

$$Y = 1.893 + .0029(\text{TIME}) - .0279(\text{TEMP}) - .0004(\text{TIME} \times \text{TEMP})$$

Table 12
Ability Decrement Calculations

- INTELLECTIVE ABILITY

$$Y = .923 + .002(\text{TIME}) - .0054(\text{TEMP}) - .0008(\text{TIME} \times \text{TEMP})$$

$$\text{DECREMENT} = \begin{cases} \dfrac{3}{10^{Y}} & \text{IF } Y >= -1 \\ 28.33 - 1.67Y & \text{IF } Y < -1 \end{cases}$$

- PERCEPTUAL-MOTOR ABILITY

$$Y = 1.893 + .0029(\text{TIME}) - .0279(\text{TEMP}) - .0004(\text{TIME} \times \text{TEMP})$$

$$\text{DECREMENT} = \begin{cases} \dfrac{3}{10^{Y}} & \text{IF } Y >= -1 \\ 25 - 5Y & \text{IF } Y < -1 \end{cases}$$

Figure 5 illustrates the logic flow that was used to calculate ability decrements as a function of heat stress in maintenance.

During a MAPPS model application, the user designates the variables which would impact performance by selecting 'HEAT STRESS' as a variable to incorporate into the model runs. Over time, heat stress would have an accumulative effect on the ability level (and other workload indications) of maintenance performance. The MAPPS model calculates the ability decrement of individuals using the heat decrement subroutine (depending upon the kind of task being simulated at that point in the model (i.e., intellective or perceptual-motor tasks in the current example).

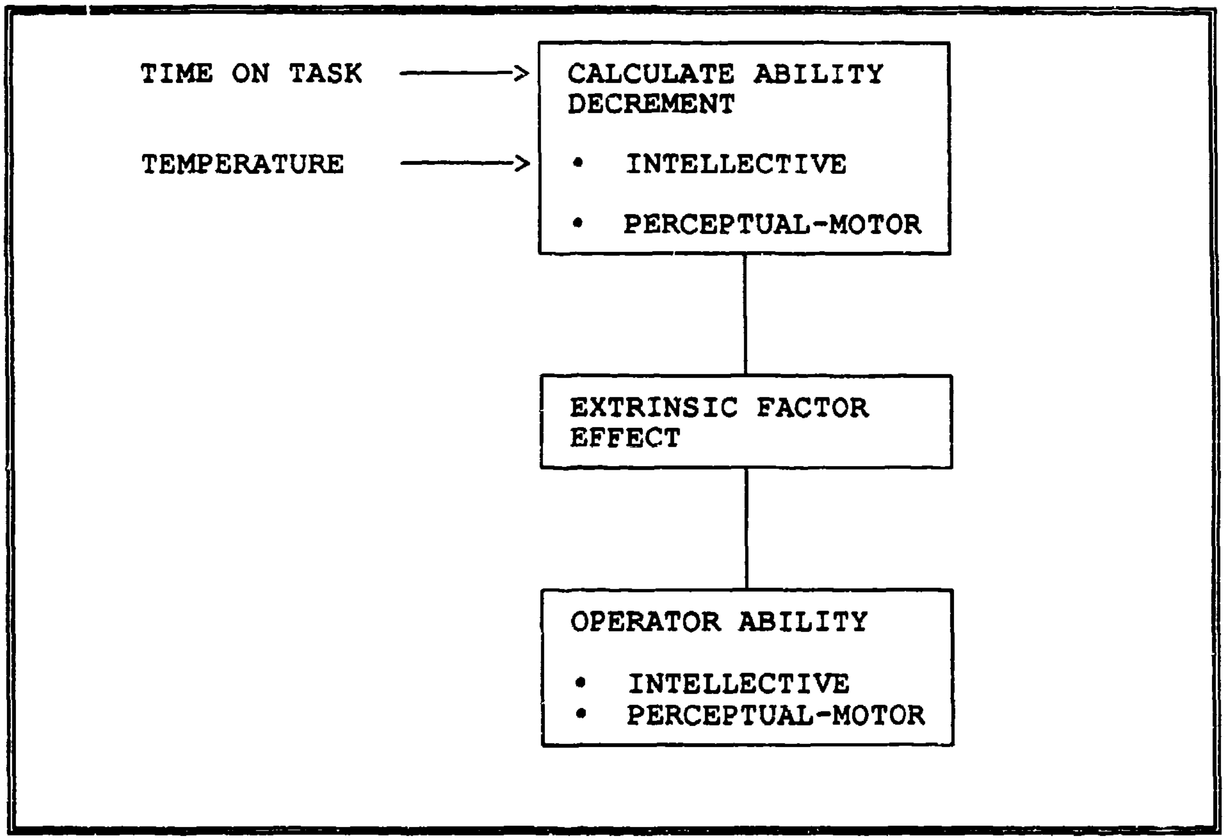


Figure 5. Ability decrement for heat stress logic flow.

In the current context, it is clear that successful prediction of the effects of heat stress on performance (defined in this instance in terms of intellective and perceptual-motor ability decrements), depends upon the very same criteria of data adequacy that were used in evaluating the learning data gleaned from the research literature for this study. That is:

- The data required to support the development of the isodecrement curves for the effects of heat stress on performance obviously existed in sufficient quantities (16 cognitive studies, 53 reaction time and cognitive studies, 7 vigilance studies, 14 tracking studies, and 48 perceptual-motor studies).

- In addition, the same (or similar) independent and dependent variables and measures were applied, as can be observed from scanning the scales around which the isodecrement functions contained in Figure 4 are based.

- While the sample size of each individual study are not reported, the pooling of data across like studies has the effect of increasing the sample size and provides added assurance of stability in the reported isodecrement relationships.

- Relevance to the target population is an open question, since the original subject samples are not described. However, performance of cognitive and perceptual-motor tasks under conditions of heat stress are clearly context-relevant taxa (i.e., tasks) to either the nuclear power plant maintenance or military environments.

- Finally, the accumulation of the findings on such a discrete topic (heat stress effects on performance) from over 100 separate research articles certainly underscores a position that the reported literature is comprehensive with respect to coverage of human behavior attendant to that taxa.

In sum, it can be categorically stated that the data used by Siegel et al. (1982) to construct their heat stress/ability decrements algorithms were perfectly suited to their use in the algorithms.

Data of a similar caliber were not found, as discussed in the opening section of this chapter. Were this situation to change (such as in the situation that new data relating the effects of training variables to performance were collected), then learning curve fitting techniques or the isoperformance methodology developed by Kennedy et al. (1988a, 1988b) (described in Chapter 2) might well be successfully applied to these kinds of analyses.

Is a Stand Alone Quantitative Model Feasible?

We can conclude based on the prior discussion that even though an adequate theoretical basis for the development of a HRGM probably exists, the extant data reported in the behavioral and social science literature do not exist in adequate quantities to support the development of a completely stand alone, quantitative model as set out in this project's initial goals. This then begs the question as to what might be done with the data collected (and reported as Appendix A). It is recommended that the data shown in Appendix A be considered for use in the development of algorithms to be incorporated into an appropriate human performance simulation model. The extent to which this is feasible, depends upon: (a) model requirements, and (b) the degree to which model users are concerned with error of prediction and/or have the means to build statistical control over such error into a simulation model. While the same limitations regarding the utility of the data apply to a human performance simulation model as they do to a stand alone quantitative model, the simulation model may be less affected by these sources of error than a stand alone model. As a consequence, the criteria and the results of our evaluation of these criteria apply to the potential application of the data contained in Appendix A as well. The following list summarizes the most serious concerns which potentially affect the utility of the reported data within a computer simulation model:

- generalizability of the data and learning curve relationships to other contexts,
- stability of the relationships given the available sample size; representativeness of the relationships contained in the data to that in each taxa at large; and relevance to the military (tasks and training), and
- usability of the data given the amount of data reported in the literature.

Conclusions and Recommendations

This study, like previous studies (e.g., Lane, 1986b; Kennedy et al., 1988a; 1988b; Spears, 1983), has come to similar conclusions regarding the theory and data available to develop models of human learning and practice. That is, the theoretical basis for the development of a model (or models) using generalizable learning curves appears acceptable. What is missing is a suitable literature and data base on which to establish such a model. Therefore, a concerted data collection effort is required in order to establish: (a) the training and performance variables of interest, (b) the performance measures and treatment groups required to collect appropriate data on these variables, and (c) the collection, reduction, and storage of the data identified with each of these steps. Once collected, the development of a HRGM can become a relatively rapid exercise.

If said data were to be collected as recommended, additional research would be required to address known issues attendant to the prediction of the effects of human learning and practice. These include

- measurement and prediction of military tasks containing multi-component operations,

- methodology for increasing the generalizability of learning curve relationships (e.g., use of individual versus group data; methodology for increasing the comparability of independent and dependent variables and measures), and

- methodology for quantifying the effects of moderator variables (e.g., aptitude, job aids, equipment design features) that mitigate the effects of human learning and practice on job performance.

Finally, it is recommended that the data contained in Appendix A be examined for potential utility in developing algorithms for use in a human performance simulation model.

References

Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. Sex Roles, 20, 327-344.

Basadur, M., Graen, G.B., & Green, S.G. (1982). Training in creative problem solving: Effects on ideation and problem finding and solving in an industrial research organization. Organizational Behavior and Human Performance, 30, 41-70.

Berliner, D., Angell, D., & Shearer, J. (1964). Behaviors, measures, and instruments for performance evaluation in simulated environments. Paper presented at a symposium and workshop on the quantification of human performance. Albuquerque, NM.

Boff, K.R., Kaufman, L., & Thomas, J.P. (1986). Handbook of perception and human performance: Volume II cognitive processes and performance. New York: Wiley.

Boff, K.R., Kaufman, L., & Thomas, J.P. (1986). Handbook of perception and human performance: Volume I Sensory processes and perception. New York: Wiley.

Boreham, N.C. (1985). Transfer of training in the generation of diagnostic hypotheses: The effect of lowering fidelity of simulation. British Journal of Educational Psychology, 55, 213-223.

Bownas, D.A., & Cooper, A. (1978). Review of reported behavior and performance taxonomies (Institute Report #17). Minneapolis, MN: Personnel Decisions Research Institute.

Briggs, G.E., & Naylor, J.C. (1965). Experiments on team training in a CIC environment (NAVTRADEVCEN 1327-1). Columbus, OH: Human Performance Center (DTIC No. AD-608309).

Card, S.K., English, W.K., and Burr, B.J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. Ergonomics, 21, 601-613.

Cockrell, J.T. (1979). Effective training for target identification under degraded conditions (ARI Technical Paper 358). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A071 085)

Cohen, J. (1977). Statistical power analysis for the behavior sciences. New York: Academic Press.

Eberts, R.E. (1987). Internal models, tracking strategies, and dual-task performance. _Human Factors_, _29_, 407-419.

Edwards, B.J., Weyer, D.C., & Smith, B.A. (1979). _Undergraduate pilot training: Visual discrimination pretraining for landing task_ (AFHRL-TR-78-78). Brooks Air Force Base, TX: Air Force Systems Command (DTIC No. AD-A068141).

Fisk, A.D., & Lloyd, S.J. (1988). The role of stimulus-to-rule consistency in learning rapid application of spatial rules. _Human Factors_, _30_, 35-49.

Gagne, R.M., Foster, H., & Crowley, M.E. (1948). The measurement of transfer of training. _Psychological Bulletin_, _45_, 97-130.

Glass, G.V. (1977). Integrating findings: The meta-analysis of research. _Review of Research in Education_, _5_, 351-379.

Grether, W.F. (1973). Human performance at elevated environmental temperatures. _Aerospace Medicine_, _44_, 747-755.

Hays, W. (1973). _Statistics for the social sciences_. New York: Holt, Rinehart, & Winston.

Hritz, R. (1979). _Behavior taxonomies and training equipment design: A literature review and general model_ (AFHRL Contract No. F33615-78-C-0019). Valencia, PA: Applied Science Associates, Inc.

Hunter, J.E., & Schmidt, F.L. (1990). _Methods of meta-analysis_. Newbury Park, CA: Sage Publications.

Jeantheau, G.G. (1971). _Handbook for training systems evaluation_ (NAVTRADEVCEN 66-C-0113-2). Orlando, FL: Naval Training Device Center.

Johnson, S.L. (1978). _Retention and transfer of training on a procedural task: interaction of training strategy and cognitive style_ (AFOSR-TR-78-1161). Bolling AFB, DC: Air Force Office of Scientific Research.

Kanfer, R., and Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. _Journal of Applied Psychology Monograph_, _74_, 657-690.

Kennedy, R.S., Jones, M.B., & Baltzley, D.R. (1988a). _Empirical demonstration of isoperformance methodology preparatory to development of an interactive expert computerized decision aid_ (ARI Research Note 88-93). (Contract No. MDA903-87-C-0603).

Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A202 439)

Kennedy, R.S., Jones, M.B., & Baltzley, D.R. (1988b). Regions of isoperformance: An interactive software package for trading off training, personnel, and equipment. (Contract No. F33615-86-C-0553). Wright-Patterson Air Force Base, OH: U.S. Air Force Systems Command.

Lane, N.E. (1986a). Issues in performance measurement for military aviation with applications to air combat maneuvering (NTSC TR-86-008). Orlando, FL: Naval Air Systems Command.

Lane, N.E. (1986b). Skill acquisition curves and military training (IDA order no. 5540). Orlando, FL: Essex Corporation.

Lintern, G., Roscoe, N.R., Koonce, J.M., & Segal, L.D. (1990). Transfer of landing skills in beginning flight training. Human Factors, 32, 319-327.

Malloy, T.E., Mitchell, C., & Gordon, O.E. (1987). Training cognitive strategies underlying intelligent problem solving. Perception and Motor Skills, 64, 1039-1046.

Masson, M.E. (1986). Identification of typographically transformed words: Instance-based skill acquisition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12, 479-488.

Miller, E.E. (1975). Instructional strategies using low-cost simulation for electronic maintenance (HumRRO Technical Report HumRRO-FR-WD-TX-75-20). Alexandria, VA: Human Resource Research Organization.

Morris, N.M., & Rouse, W.B. (1985). The effect of type of knowledge upon human problem solving in a process control task. IEEE, 15, 698707.

Myers, G.L., & Fisk, A.D. (1987). Training consistent task components: Application of automatic and controlled processing theory to industrial task training. Human Factors, 29, 255-268.

Newell, A., & Rosenbloom, P.S. (1981). Mechanisms of skill acquisition and the law of practices. In J.R. Anderson (Ed.), Cognitive skills and their acquisition (pp. 1-54). Hillsdale, NJ: Erlbaum.

O'Brien, L.H., Simon, R., & Swaminathan, H. (1989). Development of PER-SEVAL performance shaping functions. Alexandria, VA:

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

Osgood, C.E. 1949. The simulator paradox in human learning: A resolution. Psychological Review, 56, 132-143.

Patrick, J., and Haines, B. (1988). Training and transfer of fault-finding skill. Ergonomics, 31, 193-210.

Rabbit, P., Cumming, G., and Vyas, S. (1979). Improvement, Learning and Retention of Skill at Visual Search. The Experimental Psychology Society, 31, 441-460.

Ramsey & Morrissey, in Cooper, R.L. & Payne, R. (Eds.) (1978). Stress at work, Wiley: New York, NY.

Roberts, T.L., and Moran, T.P. (1983). The evaluation of text editors: Methodology and empirical results. Communications of the ACM, 26, 265-283.

Rogovin, M. & Frampton, G.T. (1979). Three Mile Island: A report to the Commissioners and to the public. Washington, DC: US Nuclear Regulatory Commission.

Roscoe, S.N. (1971). Incremental transfer effectiveness. Human Factors, 13, 561-567.

Roscoe, S.N. (1972). A little more on incremental transfer effectiveness. Human Factors, 14, 363-364.

Rouse, W.B. (1978). Human problem solving performance in a fault diganosis task. IEEE Transactions on Systems, Man, and Cybernetic, SMC-8, 258-271.

Sabol, M.A. (1986). Fitting learning curves with orthogonal polynomials (ARI Technical Report 726). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A181 148)

Schneider, W., & Fisk (1984). Automatic category search and its transfer. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 1-15.

Schneider, W. (1990). Training models to estimate training costs for new systems. In J.I. Elkind, S.K. Card, J. Hochberg, & B. M. Huey (Eds.), Human performance models for computer-aided engineering (pp. 215-232). New York: Academic Press.

Scott, P.G., McDaniel, W.C., & Braby, R. (1982). Improved procedures training through use of aids developed from learning guidelines (Navy Technical Report No. 113). Orlando, FL: Department of the Navy.

Sheppard, D.J. (1984). Visual and part-task manipulations for teaching simulated carrier landings (NAVTRAEQUIPCEN 81-C-0105-9). Westlake Village, CA: Canyon Research Group, Inc. (DTIC No. AD-A153622).

Siegel, A.I., Wolf, J.J., & Bartter, W.D. (1982). Nuclear power plant maintenance performance modeling program. Wayne, PA: Applied Psychological Services, Inc.

Simon, C.W., & Roscoe, S.N. (1981). Application of a multifactor approach to transfer of training research (NAVTRAEQUIPCEN 78-C-0060-6). Westlake Village, CA: Canyon Research Group (DTIC No. AD-A108499).

Simon, C.W. (1979). New research paradigm for applied experimental psychology: A system approach (Report No. CWS-04-77). Westlake Village, CA: Canyon Research Group, Inc.

Spears, W.D. (1983). Processes of skill performance: A foundation for the design and use of training equipment (NAVTRAEQUIPCEN Technical Report 78-C-0113-4). Orlando, FLA: Naval Equipment Training Center (DTIC No. AD-A136879).

Thurstone, L.L. (1930). The learning function. Journal of General Psychology, 3, 469-493.

Thurstone, L.L. (1919). The learning curve equation. Psychological Monographs, 26, 11-13.

Valverde, H. H. (1973). A review of flight simulator transfer of training studies. Human Factors, 15, 510-523.

Weitzman, D.O., Fineberg, M.L., Gade, P.A., & Compton, G.L (1979). Proficiency maintenance and assessment in an instrument flight simulator. Human Factors, 21, 701-710.

Wightman, D.C., & Sistrunk, F. (1987). Part-task training strategies in simulated carrier landing final approach training. Human Factors, 29, 245-254.

Wrisberg, C.A., & Winter, T.P. (1983). Variability of practice and the transfer of training of motor skills (ARI Technical Report 596). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A142 896)

APPENDIX A

DESCRIPTION OF RELEVANT ARTICLES

The purpose of Appendix A is to present general descriptions of the relevant publications obtained in the various literature searches. For each publication, a textual description is provided that includes the following information: bibliography, type of task examined in the study, taxonomy of task, type of data obtained in the study, subjects who participated, design of the study, type of training implemented, predictor variable(s), criterion variable(s), and abstract. Following the textual descriptions of each study are displays of the relevant data in graph and/or table format. Taken together, the descriptions in Appendix A should provide the reader with a means of rapidly obtaining pertinent general and specific information contained in the studies.

DESCRIPTIONS OF RELEVANT PUBLICATIONS

Basadur, M., Graen, G.B., & Green, S.G. (1982). Training in creative problem solving: Effects on ideation and problem finding and solving in an industrial research organization. Organizational Behavior and Human Performance, 30, 41-70.

Task: Creative problem solving

Taxa: Problem solving

Type of Data: Group

Subjects: 45 engineers

Design: Simple measurement, no repeated measures

Training:

1. Subjects received either a treatment consisting of training in a two day intensive creative problem solving condition, they received a placebo treatment, or they received no treatment at all.

Independent Variable(s):

1. Experiential training in creative problem solving

Dependent Measure(s):

1. Self-report attitudinal/ideation questionnaire
2. Product planning notes made by subject verbally via tape recorder
3. Recorded self-reported interviews detailing subject's observations of on the job creative problem solving performance in industrial engineering organization

Abstract: This study examined the effects of multistage creative problem solving on attitude and behaviors of subjects assessed immediately following training and two weeks after returning to work. The training addressed three critical stages: problem finding, problem solving, and solution implementation. The findings suggest that the creative problem solving training resulted in significant and measurable effects both directly following training and two weeks after returning to work.

Basadur et al, 1982 (continued)

INTERCORRELATIONS OF DERIVED SCORES REPRESENTING THE SIX DIMENSIONS[a]

	Preference for ideation in		Practice of ideation in			
	Problem finding	Problem solving	Problem finding	Problem solving	Problem-finding performance	Problem-solving performance
Preference for ideation in problem finding	.55					
Preference for ideation in problem solving	.03	.60				
Practice of ideation in problem finding	.26	.09	.94			
Practice of ideation in problem solving	.18	.39	.32	.87		
Problem-finding performance	.30	.37	.33	.49	.83	
Problem-solving performance	.04	.35	.21	.35	.22	.73

[a] Cronbach alphas have been inserted along the diagonal.

TRAINING IN CREATIVE PROBLEM SOLVING

Basadur et al, 1982 (continued)

PREFERENCE FOR IDEATION IN PROBLEM SOLVING AS MEASURED BY QUESTIONNAIRE

Measures[a]	First Reaction to New Unusual Product Ideas	Total Reaction to a New Unusual Product Idea	Deferral of Critical Judgment
Group means			
1. Experimental	5.5	6.0	12.6
2. Placebo	4.5	3.8*	10.9*
3. Untreated	4.3*	4.0	9.4**
Tests of statistical significance (α = .05)			
Group 1 vs 2	ns (p = .073)	*p = .020	*p = .044
Group 1 vs 3	*p = .031	ns (p = .062)	**p = .001
Group 2 vs 3	ns	ns	ns
$F(2,42)$	3.6*	3.8*	6.7**

[a] All measures taken immediately after training/placebo/nonplacebo.
* $p \leq .05$.
** $p \leq .01$.

PRACTICE OF IDEATION IN PROBLEM FINDING AS MEASURED BY TAPE-RECORDED, "THINKING-OUT-LOUD" TASK DURING AN INDIVIDUAL DIVERGENT THINKING TASK

Measures[a]	No. of Negative Judgments Made per Minute	Amount of Time (in seconds) Spent in Negative Evaluation
Group means		
1. Experimental	.06	2.3
2. Placebo	.34**	11.9*
3. Untreated	.36**	17.3**
Tests of statistical significance (α = .05)		
Group 1 vs 2	**p = .004	*p = .017
Group 1 vs 3	**p = .013	**p = .001
Group 2 vs 3	ns	ns
$F(2,42)$	5.4**	4.5*

[a] All measures taken immediately after training/placebo/nonplacebo.
* $p \leq .05$.
** $p \leq .01$.

PRACTICE OF IDEATION IN PROBLEM SOLVING
METHOD OF MEASUREMENT

	Questionnaire				Interview			
	Change Noticed in Being				Change Noticed in Being			
	More Openminded to New Ideas and Approaches		More Openminded to New Ideas and Approaches		Less Likely to Jump to Conclusions as to what Is the Real Problem		More Likely to Pause to Try Unusual or Creative Approaches to Solving Problems	
Measures[a]	By self	By others	By self	By others	By self	By others	By self	By others
Group means								
1. Experimental	+3.6	+1.5	+3.0	+0.63	+2.1	+.52	+1.8	+0.67
2. Placebo	+1.8**	+1.0*	+0.7**	+0.30*	+0.1**	+.19	+0.6	+0.07**
3. Untreated	+1.2**	+0.9*	+1.2**	+0.17**	+0.4**	+.04**	+0.2*	+0.00**
Tests of statistical significance (α = .05)								
Group 1 vs 2	**p = .001	*p = .045	**p < .001	*p = .050	**p < .001	ns(p = .067)	ns(p = .076)	**p = .002
Group 1 vs 3	**p = .001	*p = .022	**p < .001	**p = .007	**p < .001	**p = .005	*p = .026	**p = .001
Group 2 vs 3	ns	ns	ns	ns	ns	ns	ns	ns
$F(2,42)$	14.4**	3.5*	13.3**	5.2**	20.4**	5.1**	3.5*	13.5*

[a] All measures taken 2 weeks after return to work.
* $p \leq .05$.
** $p \leq .01$.

PROBLEM-FINDING PERFORMANCE

Measures*	Method of measurement: Tape-recorded, thinking-out-loud" task: Wishes generated for a new product of the future — Creative quality (based on Jackson & Messick's 4 criteria of a creative product)	Quantity (total number of wishes generated)	Amount of time (in seconds) spent in divergent, problem-finding thought during task	Questionnaire — Number of different problem definitions developed prior to choosing one as best
Group means				
1. Experimental	1.8	24.3	177.7	8 4
2. Placebo	1.1*	13.9*	121.4**	2.6**
3. Untreated	1.0**	12.2 *	120.6**	2.5**
Tests of statistical significance (*α = .05)				
Group 1 vs 2	*p = .015	*p = .020	**p = .009	**p = .002
Group 1 vs 3	**p = .004	**p = .007	**p = .012	**p = .002
Group 2 vs 3	ns	ns	ns	ns
$F(2,42)$	5.4**	6.3**	4.8**	12.0**

* All measures taken immediately after training/placebo/nonplacebo.
*$p \leq .01$.
**$p \leq .05$.

PROPORTION OF INDIVIDUAL MEASURE COMPARISONS SUPPORTING EACH HYPOTHESIS[a]

	In an applied research setting, given a sample that has a relatively low ideation tendency, training in a "complete process of problem solving" emphasizing the ideation–evaluation process in all stages (Fig. 1) will lead to an increase in					
	Preference for ideation in problem finding (H_{1A})	Preference for ideation in problem solving (H_{1B})	Practice of ideation in problem finding (H_{2A})	Practice of ideation in problem solving (H_{2B})	Problem-finding performance (H_{3A})	Problem-solving performance (H_{3B})
Immediately after training	0/6 (One method)	4/6 (One method)	4/4 (One method)	NA	8/8 (Two methods)	NA
Two weeks after return to work setting	NA[b]	NA	NA	14/16 (Two methods)	NA	1/4 (One method)

[a] Expressed as the fraction of the total, the number of comparisons providing a significantly higher mean score vs placebo/nonplacebo at $\alpha = .05$.

[b] NA, not applicable (i.e., no measures available).

Boreham, N.C. (1985). Transfer of training in the generation of diagnostic hypotheses: The effect of lowering fidelity of simulation. British Journal of Educational Psychology, 55, 213-223.

Task: Simulated problem diagnosis

Taxa: Problem solving

Type of Data: Group

Subjects: 40 adults

Design: Repeated measures

Training:

1. Subjects were trained for either divergent or convergent hypothesis testing in problem solving.

Independent Variable(s):

1. Hypothesis testing training for simulated diagnostic tasks used in salt packaging plant

Dependent Measure(s):

1. Number of different categories of hypotheses generated during pre-test and post-test

Abstract: This study tested whether practice of diagnostic problem solving inhibits learning by forcing the learner to practice the convergent type of hypothesis typical of most diagnostic situations. Forty adults practiced on a simulation of a diagnostic task and were then tested on their ability to generate hypotheses on a new problem. The degree of convergence/divergence of the hypothesis testing was varied experimentally. Results demonstrated that subjects who practiced the more convergent type of hypothesis testing acquired less ability than those who practice the more divergent type of hypothesis testing.

PRE- AND POST-TEST SCORES FOR THE NUMBER OF CATEGORIES OF HYPOTHESIS GENERATED

	Pre-test	Post-test	Difference
Condition 1 mean	3·10	3·17	0·07
(SD)	(0·94)	(0·87)	(1·13)
Condition 2 mean	2·58	3·45	0·87
(SD)	(1·33)	(0·99)	(1·10)

NUMBER OF STUDENTS WHOSE POST-TEST SCORE SHOWED AN IMPROVEMENT ON THEIR PRE-TEST SCORE

Change from Pre-test to Post-test	Condition 1	Condition 2
Improvement	9	16
No Improvement	11	4

Fisher's exact test (two-tailed) $P < 0{\cdot}05$

Briggs, G.E., & Naylor, J.C. (1965). Experiments on team training in a CIC environment (NAVTRADEVCEN 1327-1). Columbus, Ohio: Human Performance Center (DTIC No. AD-608309).

Task: Simulated radar-controlled aerial intercept

Taxa: Information processing/fine motor

Type of Data: Group

Subjects: 24 Radar Controllers

Design: Factorial, 2(task complexity) x 2(task organization) x 2(replacement training)

Training:

1. Subjects were trained on a simulated radar-controlled aerial intercept task either independently or interactively.

2. Subjects were given either a relatively complex or a relatively simple task.

3. Subjects were either given additional training sessions or fewer training sessions.

Independent Variable(s):

1. Interaction of team members in training for radar control task
2. Complexity of radar control task
3. Length training for radar control

Dependent Measure(s):

1. Efficiency score composed of number of successful intercepts, time, and fuel consumed

Abstract: Two experiments designed to study team training in a Combat Information Center (CIC) type environment were conducted. The transfer task of both experiments required the two team members to coordinate their radar-controlled air intercepts. Training task fidelity was varied in terms of input and output features of the task environment in a first experiment and in terms of task definition in a second experiment. Both input and output fidelity affected the ability to coordinate at transfer. Coordination performance at transfer tended to be directly related to the emphasis placed on coordination skills during training.

Percentage of Subjects at Each College Level for the Three Experiments

Experiment	Years in School 1	2	3	4	5
I	25.7	23.5	23.5	19.1	8.0
II	30.0	18.7	15.1	20.5	15.1
III	37.5	31.2	20.8	6.2	4.1

Analysis of Efficiency Data for Sessions 1 and 2
of Experiment III

Source	df	MS	F
Transfer Organization (TrO)	1	385,193,984	1.02
Training Organization (TO)	1	3,667,968	-
TO x TrO	1	728,579,072	1.92
Teams within groups (Ts/G)	20	378,741,424	
Sessions (S)	1	5,469,124,608	21.38**
S x TrO	1	493,853,184	1.93
S x TO	1	100,352	-
S x TO x TrO	1	657,139,712	2.57
S x Ts/G	20	255,743,856	

** $p < .01$

Average Efficiency Data for Sessions 1 and 2 of Experiment II
(Group numbers indicated in parentheses)

Transfer Task	Training Task					
	Low Fidelity			High Fidelity		
	Interact.	Indep.	Sums	Interact.	Indep.	Sums
Interaction	19,193 (8)	10,922 (6)	30,115	16,695 (7)	9,213 (5)	25,908
Independent	7,872 (4)	11,175 (2)	19,047	11,650 (3)	9,975 (1)	21,625

Main Effects (Sums)

Training Task Fidelity		Task Organization			
		Training Task		Transfer Task	
Low	High	Interact.	Indep.	Interact.	Indep.
49,162	47,733	55,410	41,275	56,023	40,672

Briggs et al., 1965 (continued)

Analyses of the Efficiency Data of Experiment II

Source	df	Sessions 1 & 2 MS	F	Sessions 3 & 4 MS	F
Fidelity (F)	1	4,642,176	-	2,329,088	-
Training Organization (TO)	1	349,123,456	4.23*	697,792	-
Transfer Organization (TrO)	1	412,416,334	5.00*	78,733,184	2.18
F x TO	1	30,711,936	-	9,247,808	-
F x TrO	1	80,538,240	-	60,456,000	1.67
TO x TrO	1	528,721,792	6.41*	64,082,752	1.77
F x TO TrO	1	58,196,864	-	3,849,152	-
Teams within Groups (Ts/G)	48	82,470,515		36,175,124	
Sessions (S)	1	684,934,400	34.84**	26,712,832	2.52
S x F	1	684,544	-	589,696	-
S x TO	1	95,744	-	12,496,368	1.19
S X TrO	1	54,917,888	2.79	6,565,824	-
S x F x TO	1	7,382,016	-	19,252,056	1.82
S x F x TrO	1	91,826,432	4.67*	5,068,160	-
S x TO x TrO	1	27,614,208	1.40	9,108,672	-
S x F x TO x TrO	1	21,376	-	16,354,240	1.54
S X Ts/G	48	19,660,482		10,605,585	

* $p < .05$
** $p < .01$

Briggs et al., 1965 (continued)

The Interaction of Sessions and Replacement Training Variables in Experiment I

Amount of Replacement Training	Transfer Session 2	Transfer Session 3
High	19,359	10,020
Low	11,826	12,147

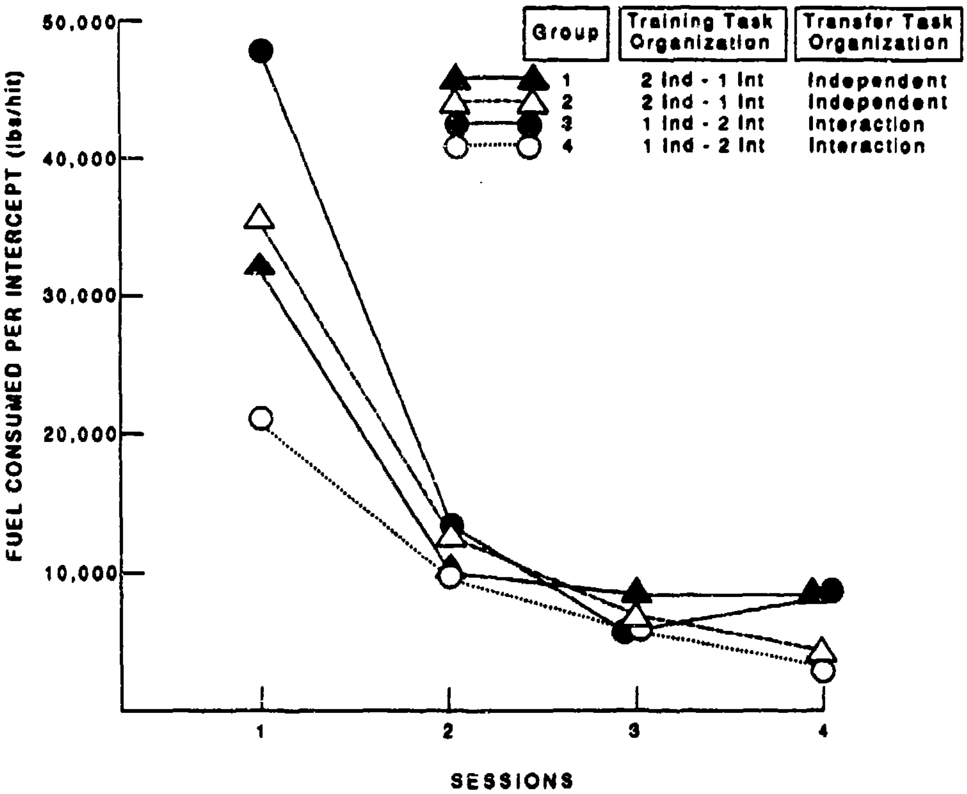
Briggs et al., 1965 (continued)

Analyses of the Efficiency Data (Fuel Consumed per Successful Interception) of Experiment I

Source	df	Sess. 1&2 F (MS)	Sess. 2&3 F (MS)	Sess. 3&4 F (MS)
Task Organization (O)	1	- (7,611)	3.64 (275,134)	7.66* (140,764)
Replacement Training (R)	1	2.68 (1,382,521)	1.54 (116,894)	- (2,705)
Task Complexity (C)	1	12.77** (6,577,565)	10.99** (829,872)	19.74** (362,400)
O x R	1	- (403,361)	- (3,894)	- (7,998)
O x C	1	- (19,066)	- (5,838)	- (7,998)
R x C	1	- (86,754)	- (5,598)	- (1.023)
O x R x C	1	- (87,643)	- (566)	- (1,158)
Teams withing Groups (Ts/G)	24	(514,764)	- (75,490)	- (18,356)
Sessions (S)	1	21.59** (11,224,069)	5.31* (325,329)	4.38* (81,647)
S x O	1	- (337,544)	- (30,238)	- (602)
S x R	1	- (49,669)	6.09* (373,372)	2.53 (47,145)
S x C	1	3.24 (1,688,088)	2.04 (125,648)	- (2,069)
S x O x C	1	- (223,235)	- (5,575)	- (2,323)
S x O x R	1	- (419,085)	1.08 (66,670)	- (8,531)
S x R x C	1	- (81,018)	- (4,214)	- (11,610)
S x O x R x C	1	- (65,620)	- (258)	- (1,742)
S x Ts/G	24	(519,763)	(61,295)	(18,611)

* $p < .05$
** $p < .01$

Briggs et al., 1965 (continued)



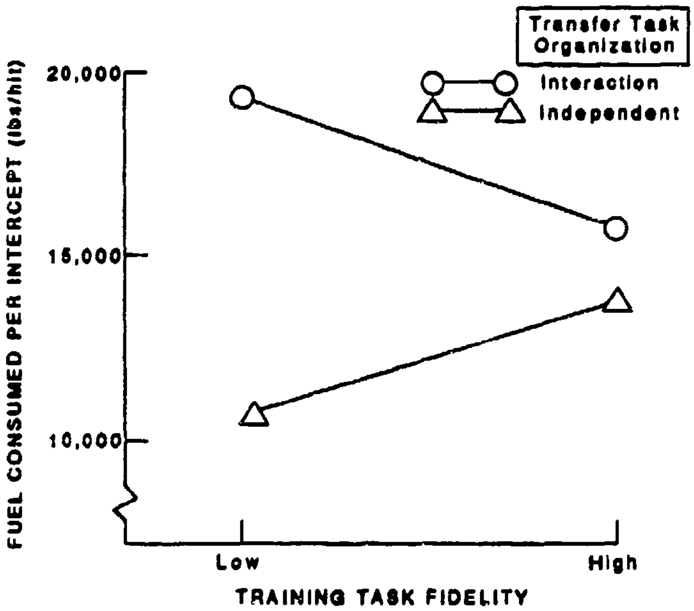
Efficiency data for Experiment III. The legend for training task organization represents the number of sessions spent in individual (Ind) and team or interaction (Int) training.

Briggs et al., 1965 (continued)

The Design of Experiment III

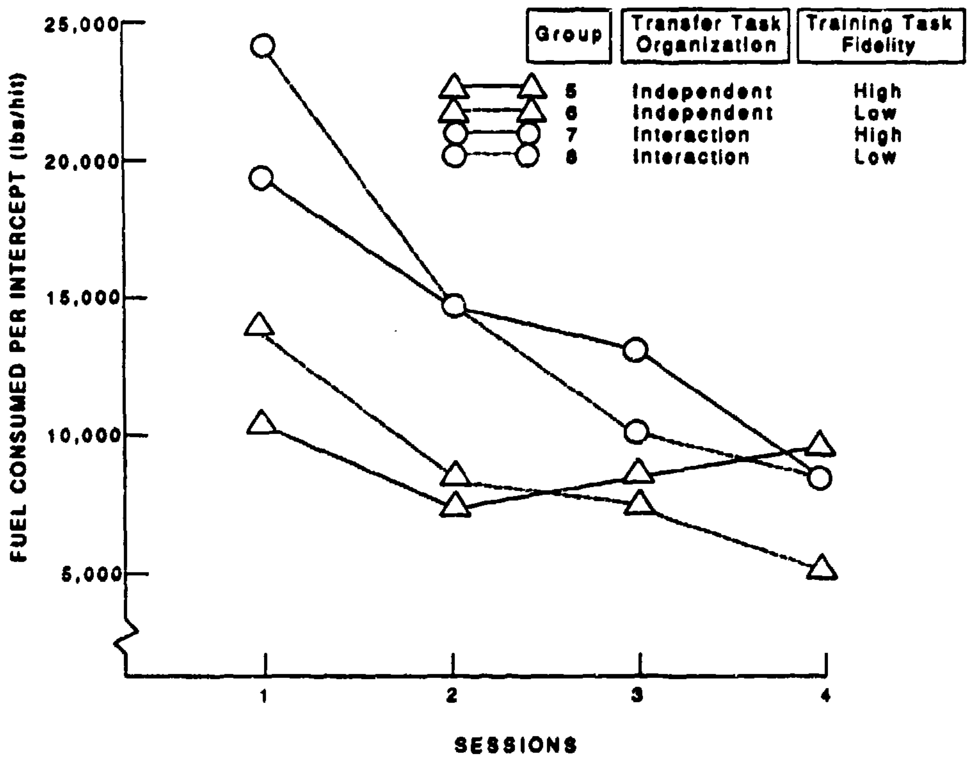
Group	Training Emphasis	Transfer Task Organization
1	2 Ind - 1 Int	Independent
2	2 Ind - 1 Int	Interaction
3	1 Ind - 2 Int	Independent
4	1 Ind - 2 Int	Interaction

Briggs et al., 1965 (continued)

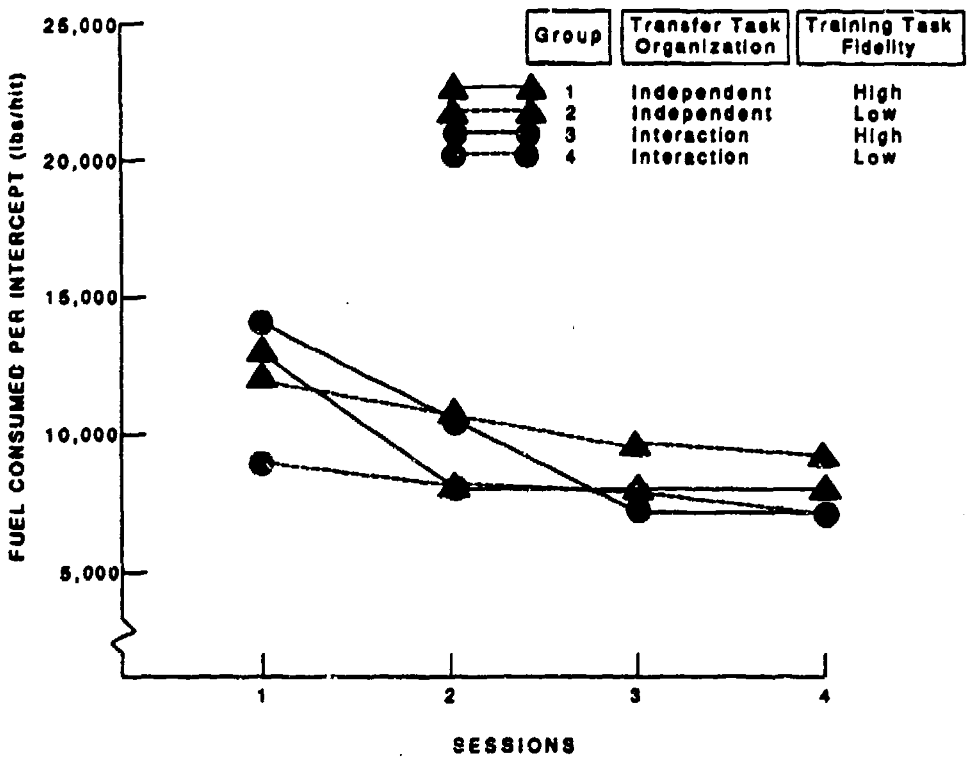


Average efficiency under the four combinations of training task fidelity and transfer task organization of Experiment II.

Briggs et al., 1965 (continued)

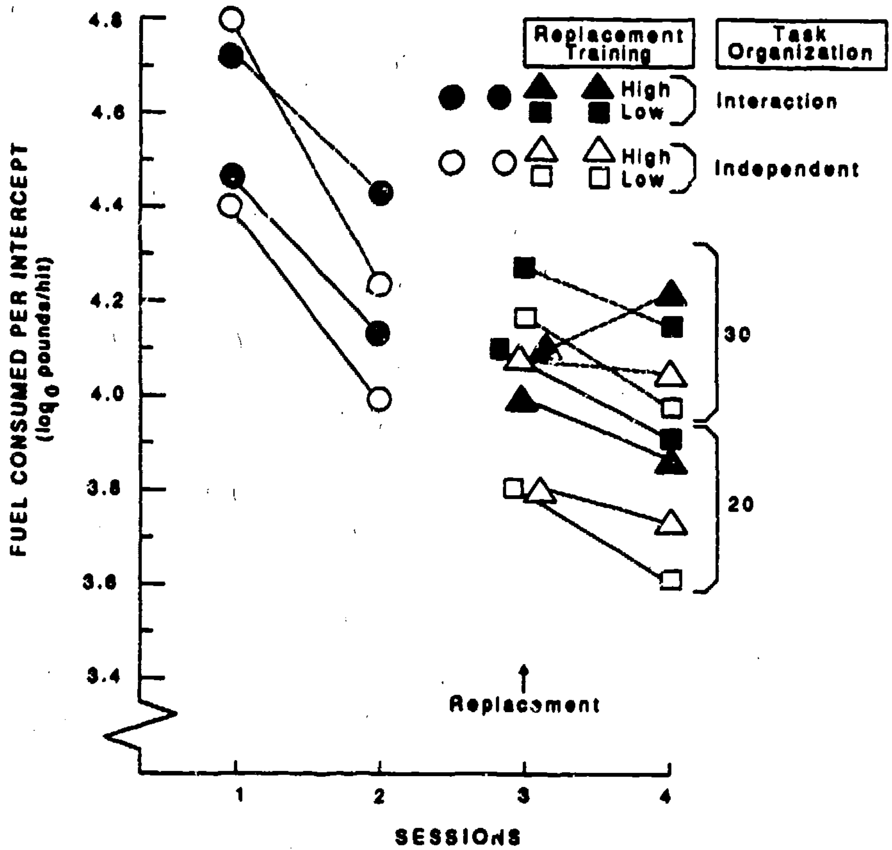


Efficiency data for the interaction organization of the operational task. Experiment III.



Efficiency data for the independent organization of the operational task. Experiment II.

Briggs et al., 1965 (continued)



System efficiency for Experiment I.

Card, S.K., English, W.K., and Burr, B.J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. Ergonomics, 21, 601-613.

Task: Text selection on a CRT display

Taxa: Fine motor/Information processing

Type of Data: Group means

Subjects: 5 undergraduate students

Design: Repeated measures

Training:

1. Subjects were trained on four different text selecting devices

2. Performance was tested during latter section of training trials

Independent Variable(s):

1. Distance from starting point to Positioning device (mouse, joystick, step keys, and text keys)

Dependent Measure(s):

1. Positioning speed
2. Homing time

Abstract: This study evaluated four devices (mouse, isometric joystick, step keys, and text keys) with respect to the speed with which they could be used to select text on a CRT display. Results demonstrated that the mouse was fastest and produced the least errors. It was shown that variations in positioning time with the mouse and joystick are accounted for by Fitts's Law. With regard to the mouse, the measured Fitts's Law slope constant is close to that found in other eye-hand tasks leading to the conclusion that positioning time with the mouse is almost the minimal achievable. With regard to key devices, positioning time was shown to be proportional to the number of keystrokes that must be typed.

Summary of Models for Positioning Time (T_{pos})

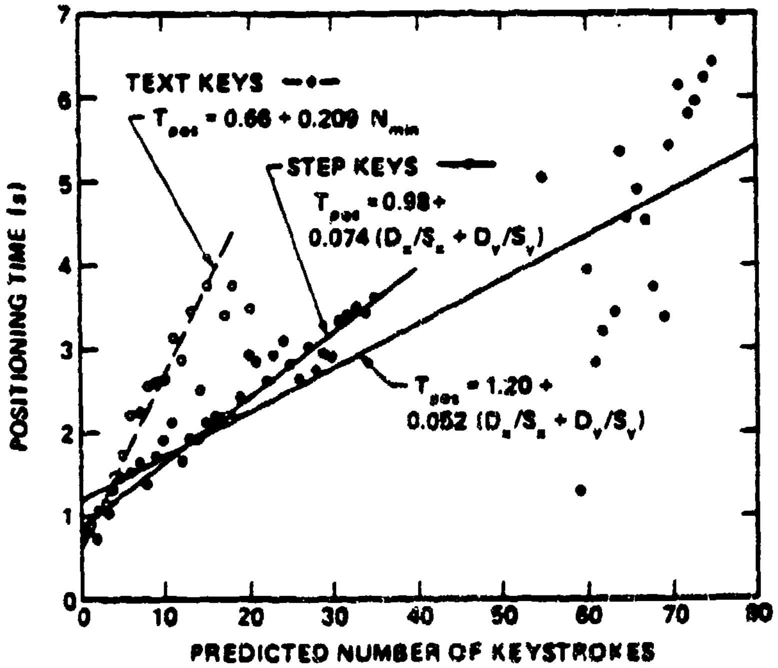
Device	Model (times in s)	s_e	R^2
Mouse	$T_{pos} = 1.03 + 0.096 \log_2 (D/S + 0.5)$	0.07	0.83
Joystick	$T_{pos} = 0.99 + 0.220 \log_2 (D/S + 0.5)$[a]	0.13	0.89
	$T_{pos} = K_d + 0.1 \log_2 (D/S + 0.5)$[b]	0.07	----
Step Keys	$T_{pos} = 1.20 + 0.052 (D_x/S_x + D_y/S_y)$[c]	0.54	0.84
	$T_{pos} = 0.98 + 0.074 (D_x/S_x + D_y/S_y)$[d]	0.18	0.95
Text Keys	$T_{pos} = 0.66 + 0.209 N_{min}$	0-24	0-89

[a]Least squares fit to all data points.

[d]Fit for number of keystrokes $D_x/S_x + D_y/S_y$) <40.

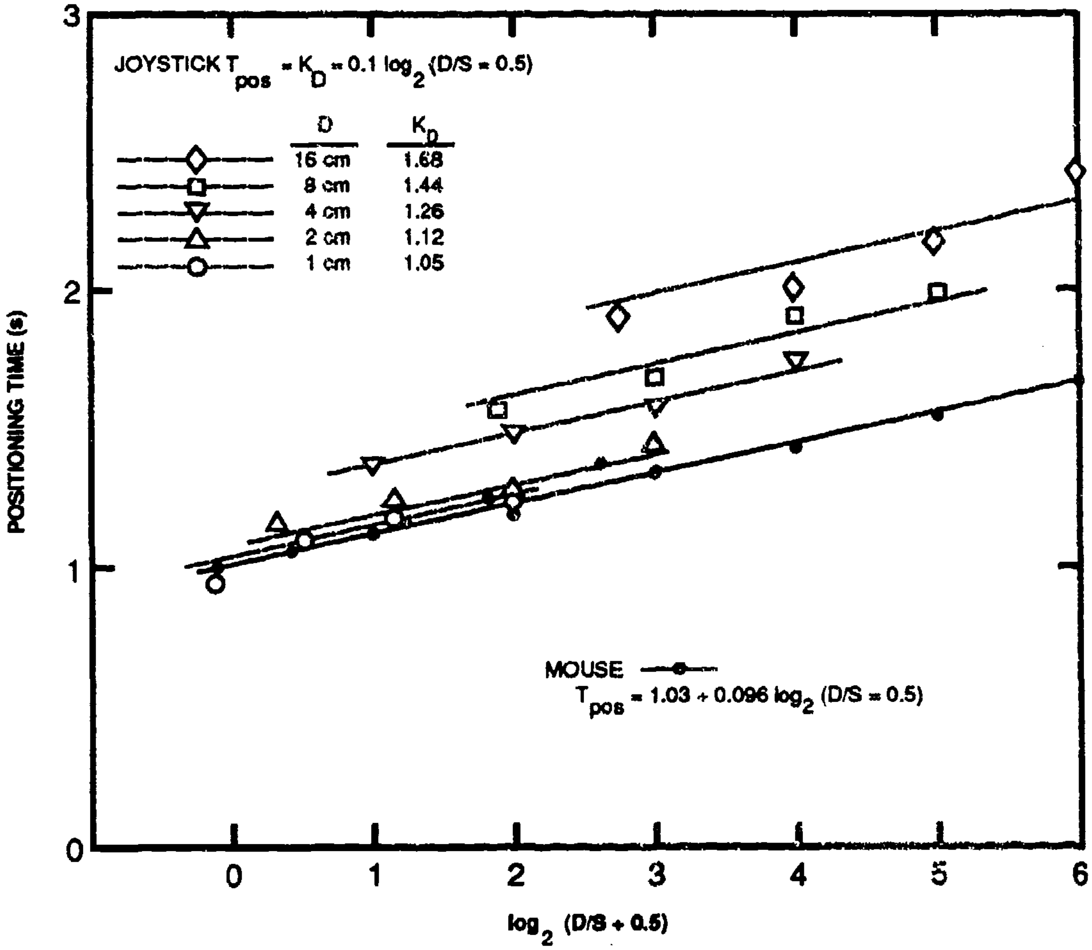
[c]Least squares fit to all data points.

[b]Fitting a separate line with slope 0.1 bits $^{-1}$ for each distance where HOME key unlikely to be used.



Positioning time for key devices as a function of predicted number of keystrokes.

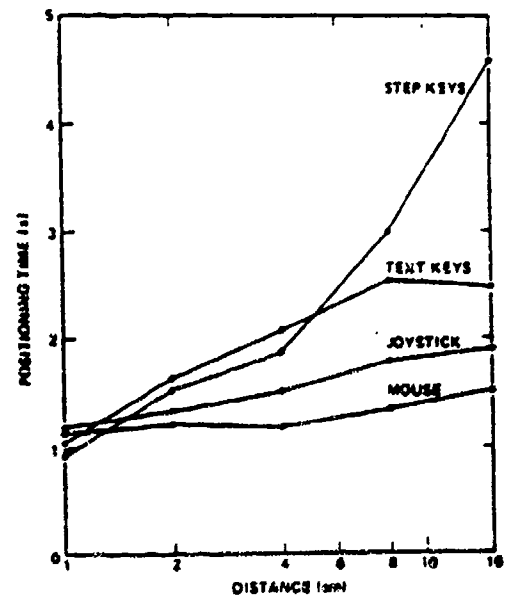
Card et al, 1978 (continued)



Card et al, 1978 (continued)

Overall Times

Device	Movement time for non-error trials (s)						Error rate	
	Homing Time		Positioning Time		Total Time			
	M	SD	M	SD	M	SD	M	SD
Mouse	0·36	0·13	1·29	0·42	1·66	0·48	5%	22%
Joystick	0·26	0·11	1·57	0·54	1·83	0·57	11%	31%
Step Keys	0·21	0·30	2·31	1·52	2·51	1·64	13%	33%
Text Keys	0·32	0·61	1·95	1·30	2·26	1·70	9%	28%



Effect of target distance on positioning time.

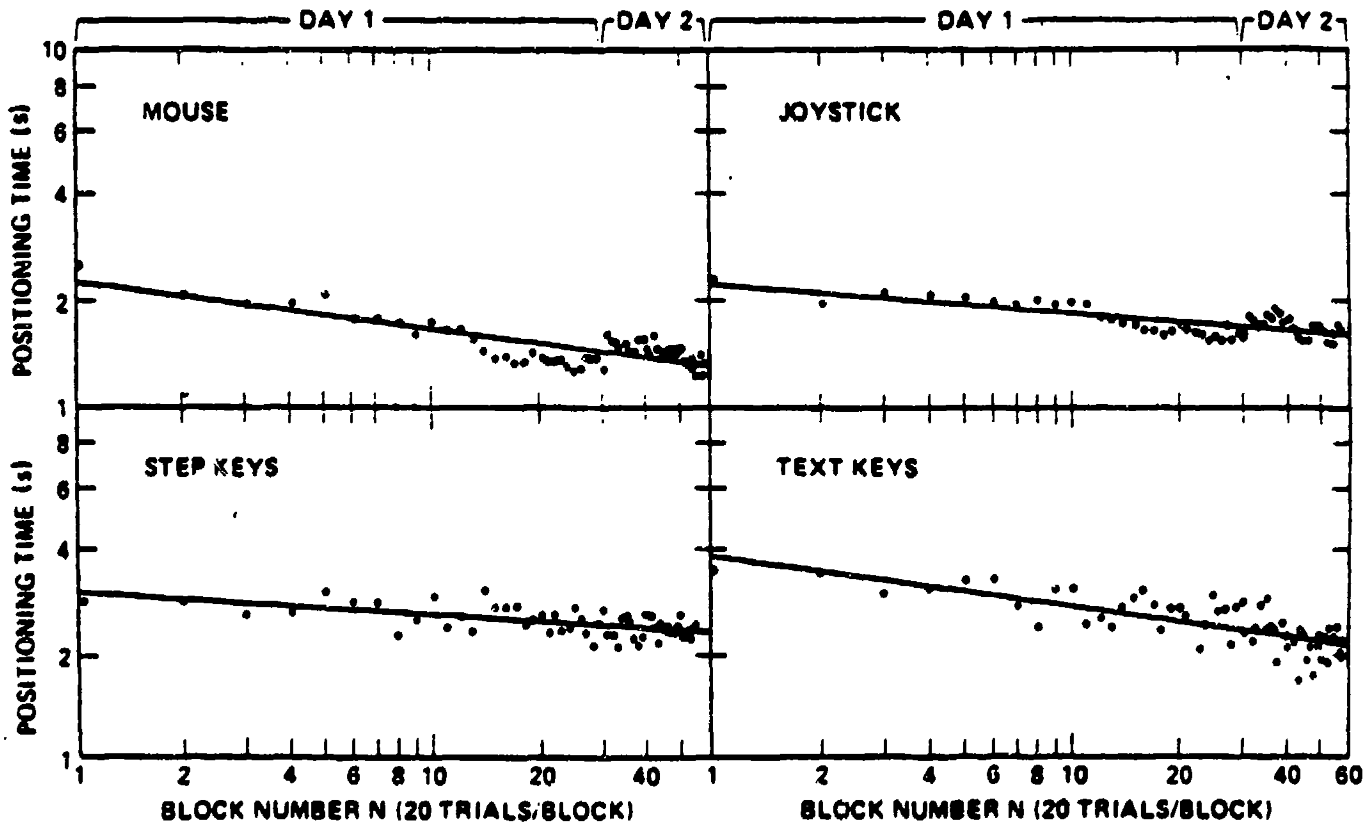
Card et al, 1978 (continued)

Learning Curve Parameters

DEVICE	T_1 (s)	x	Learning Curve Equation*	s_e (s)	R^2
Mouse	2.20	0.13	$T_N = 2.20\, N^{-0.13}$	0.12	0.66
Joystick	2.19	0.08	$T_N = 2.19\, N^{-0.08}$	0.08	0.62
Step Keys	3.03	0.07	$T_N = 3.03\, N^{-0.07}$	0.11	0.39
Text Keys	3.86	0.15	$T_N = 3.86\, N^{-0.15}$	0.16	0.61

*N is number of trial blocks. There are 20 trials in each block.

Card et al, 1978 (continued)



Learning curves for pointing devices.

Cockrell, J.T. (1979). Effective training for target identification under degraded conditions (ARI Technical Paper 358). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Task: Visual discrimination

Taxa: Visual

Type of Data: Group

Subjects: 96 subjects who had just completed basic training or who were in advanced individual training

Design: Repeated measures

Training:

1. Subjects were trained on targets of differing visibility

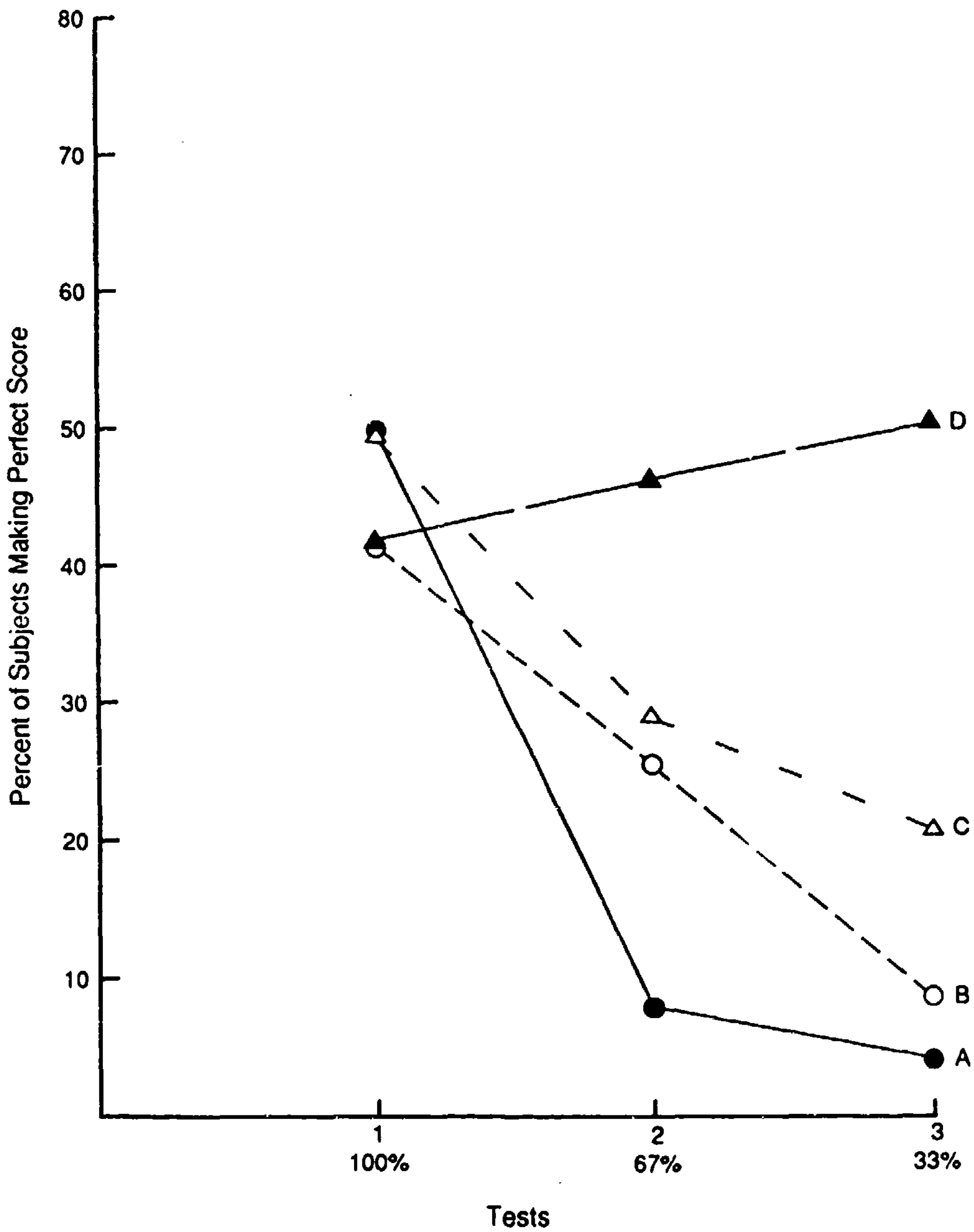
Independent Variable(s):

1. Target quality during training (degraded vs. normal)

Dependent Measure(s):

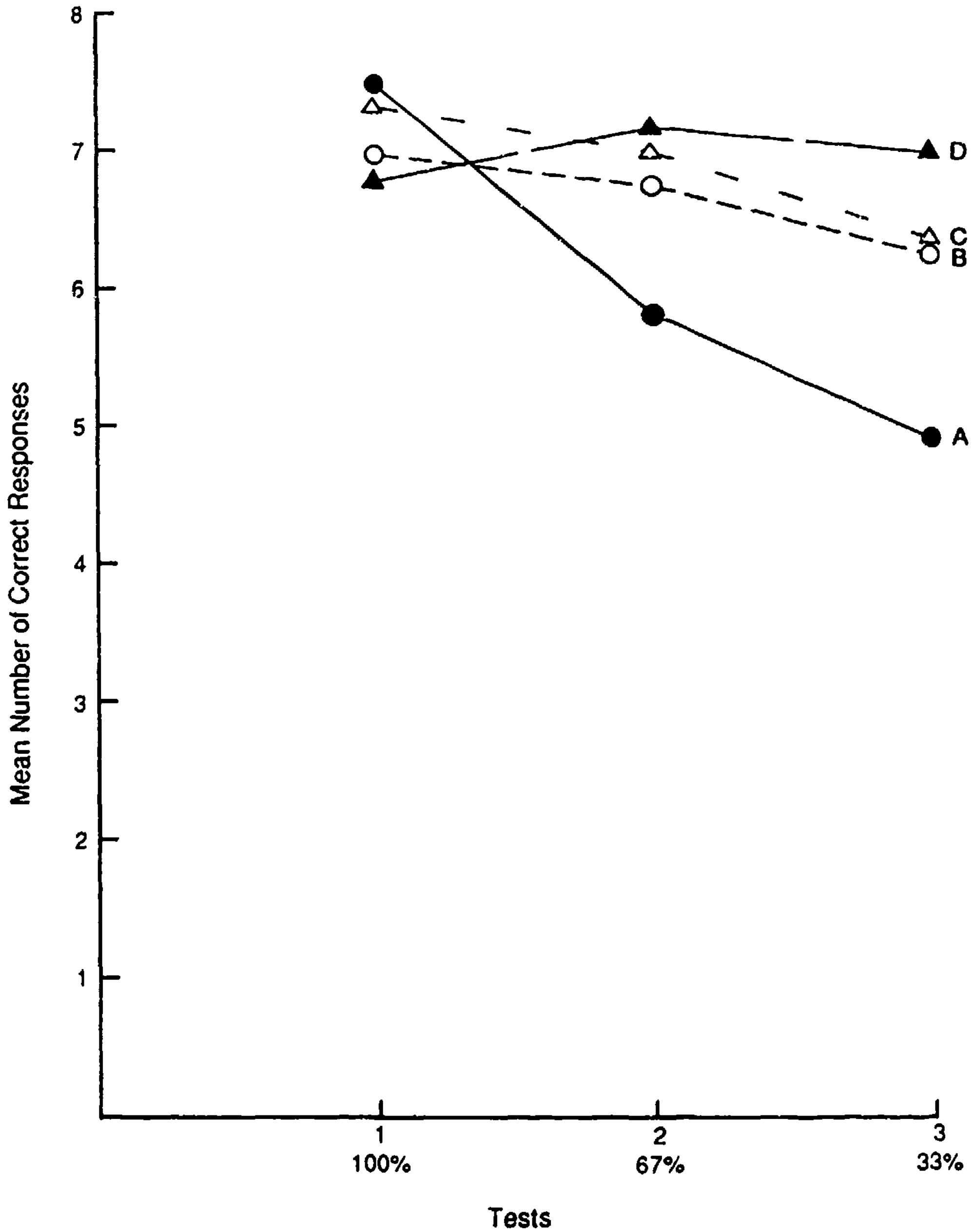
1. Mean number of correct identifications of degraded targets

Abstract: This study explored the relevance of the psychological concept of overshadowing to military target identification training. The concept would predict that the most outstanding features of the target will capture the trainees attention more than any of the other features. This suggests that soldiers will have difficulty identifying degraded targets unless they are trained to do so. This experiment forced trainees to focus attention on increasing numbers of features by occluding the outstanding features. Results demonstrated the relevance of overshadowing and suggested that most training should concentrate on degraded views.



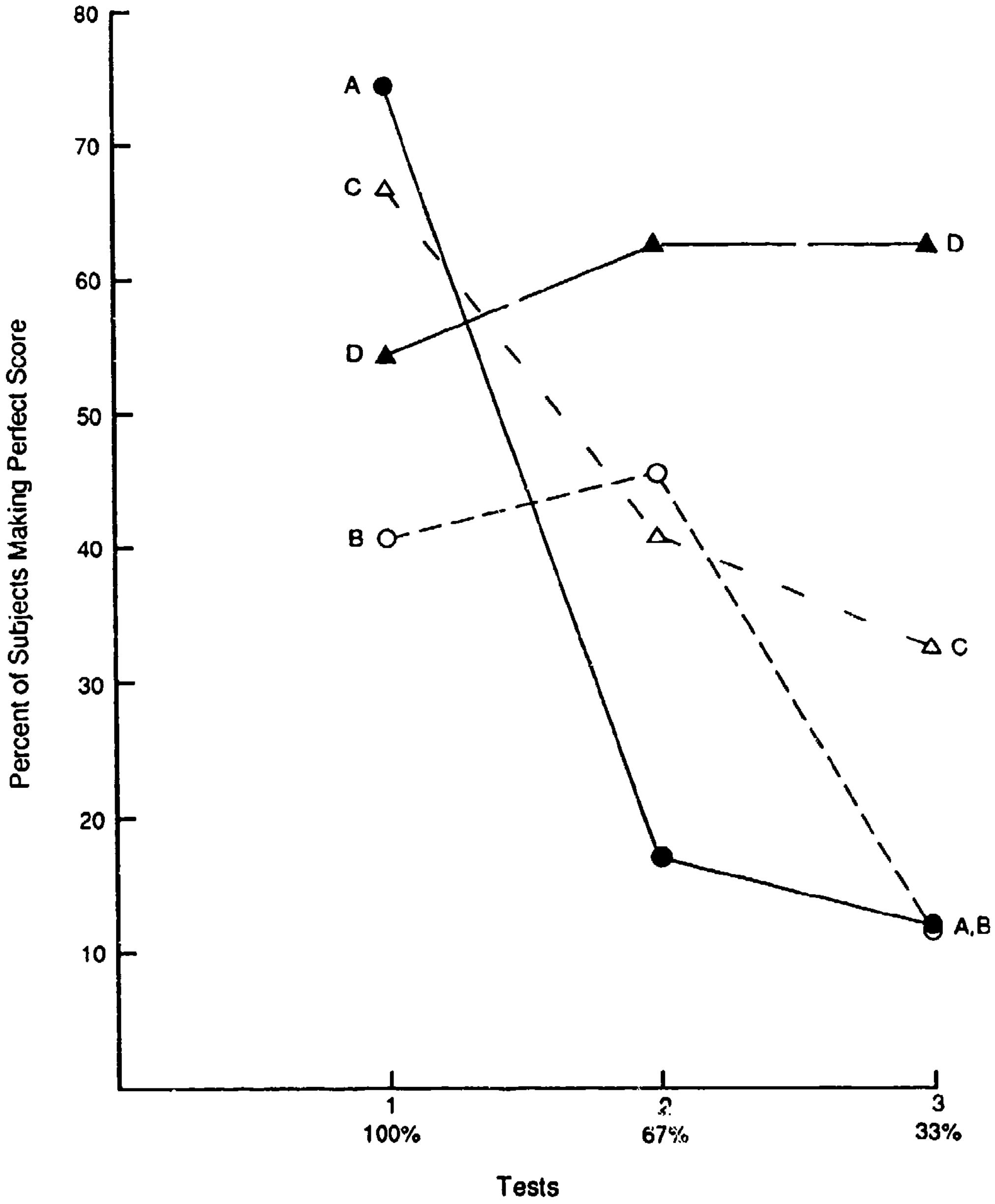
Percent of subjects making perfect score on both
main experiment test and retest.

Cockrell, 1979 (continued)

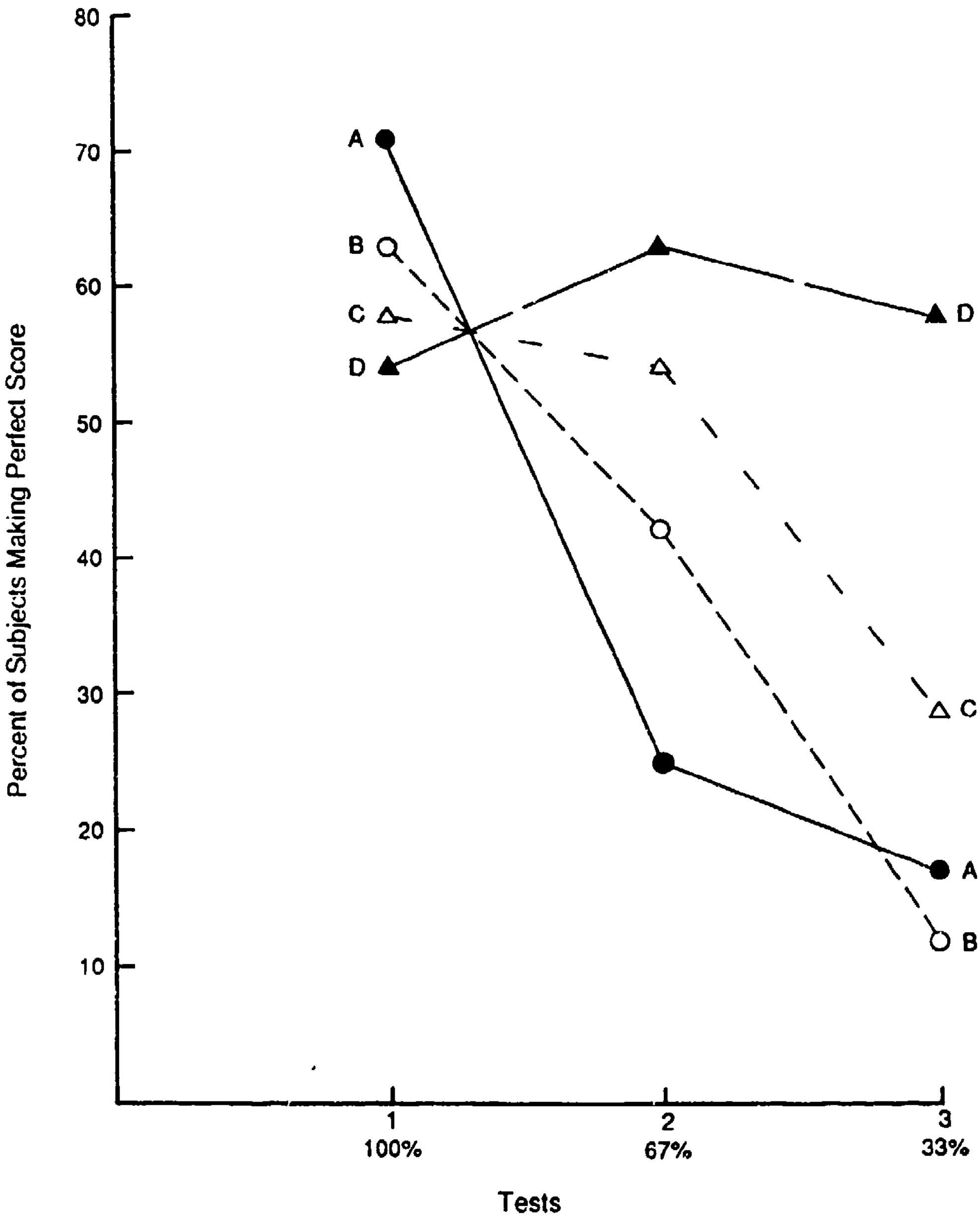


Mean number of correct responses on main experiment tests.

Cockrell, 1979 (continued)

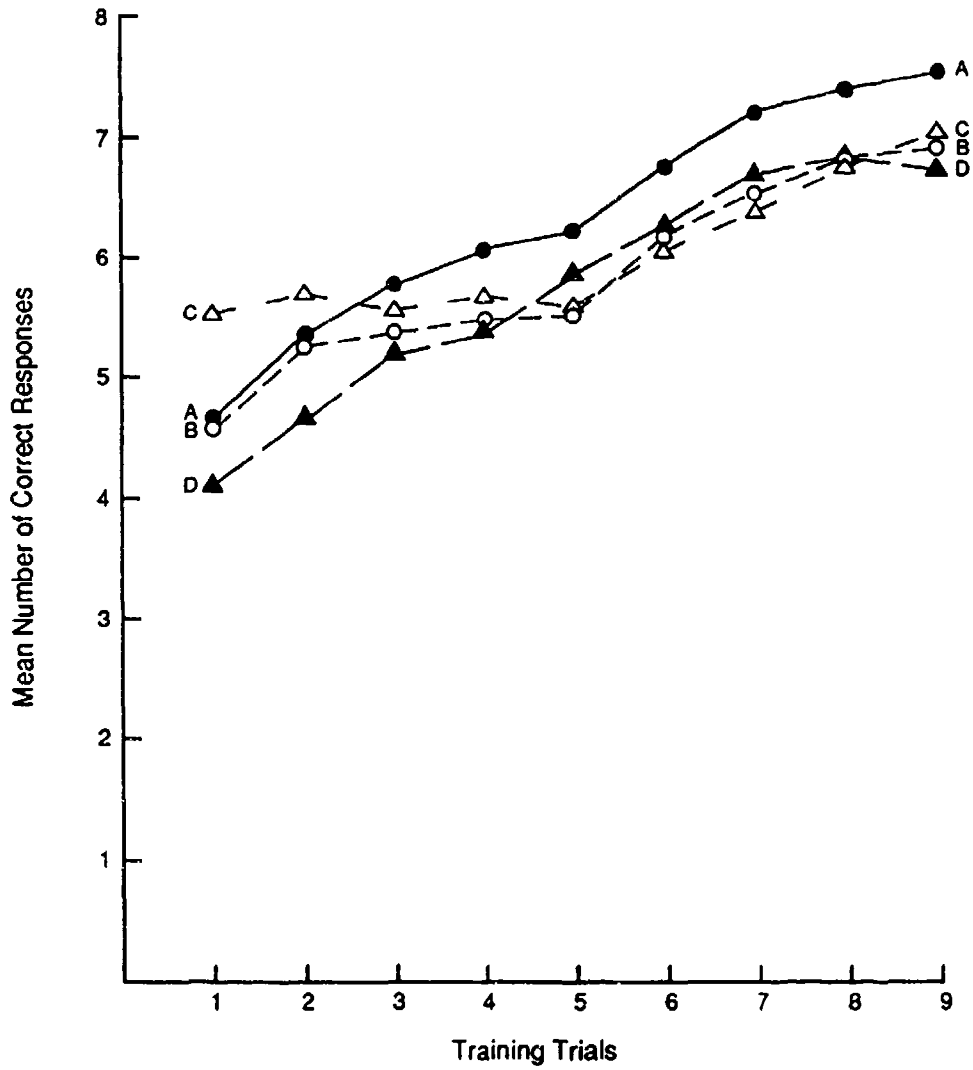


Percent of subjects making perfect score on main experiment tests.

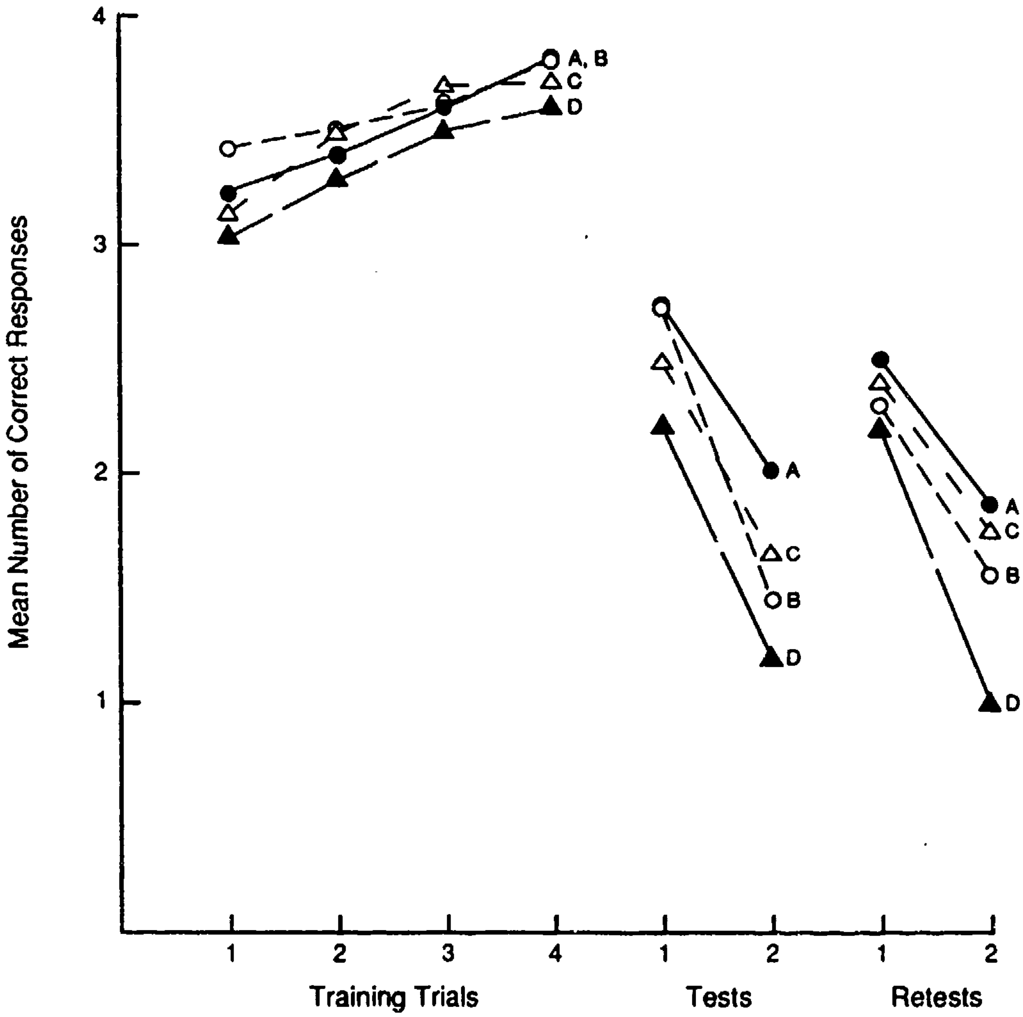


Percent of subjects making perfect score on main experiment retests.

Cockrell, 1979 (continued)



Mean number of correct responses on main experiment training trials. (Smoothed curve.)



Mean number of correct responses for each group during the warm-up phase. (Smoothed curve.)

Eberts, R.E. (1987). Internal models, tracking strategies, and dual-task performance. Human Factors, 29, 407-419.

Task: Visual tracking and auditory detection task

Taxa: Visual/Information processing

Type of Data: Group

Subjects: 24

Design: 3(parabola augmentation, point augmentation, or no augmentation) x 2(single or dual task)

Training:

1. Subjects performed either a single task (auditory or visual) or a dual-task (both auditory and visual).

2. Subjects were given either point cues, parabola cues, or no cues.

Independent Variable(s):

1. Task type (single or dual)

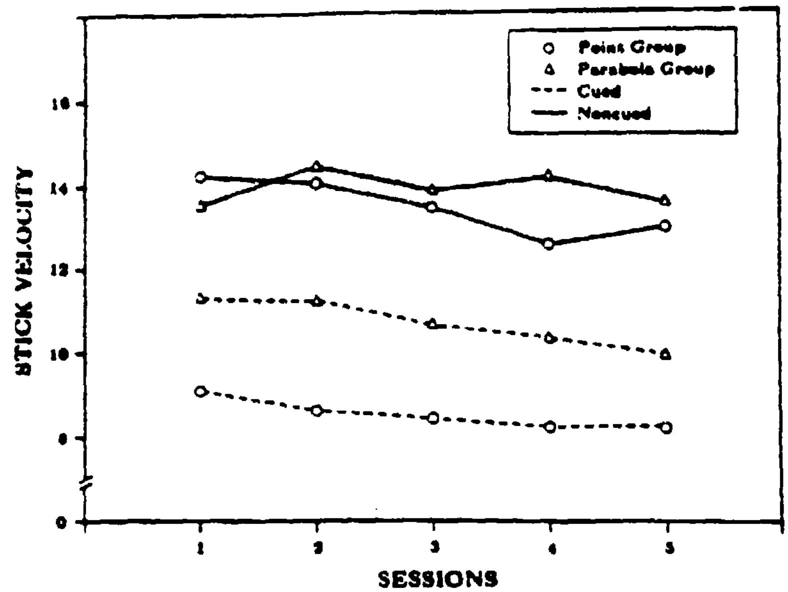
2. Augmentation type (point or parabola cues)

Dependent Measure(s):

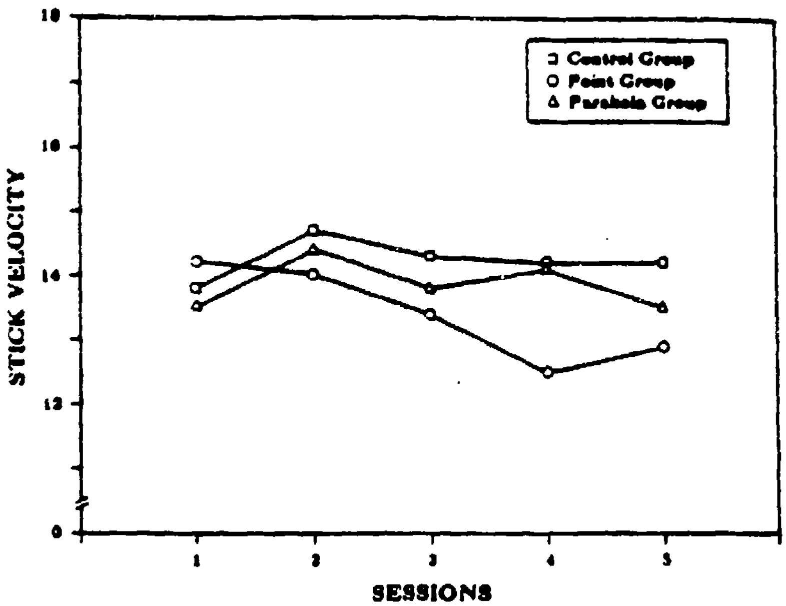
1. Root mean squared error (RMS) tracking performance

Abstract: This study investigated the effects of internal models and tracking strategies on workload. Internal models and tracking strategies were varied by providing subjects with augmenting cues. Performance of two groups provided with different types of display augmentation (parabola or point cues) was compared with that of a control group for single-task tracking training. Results indicated that both point and parabola augmentation reduced workload when displayed on cued trials. On the non-cued trials, the parabola augmentation reduced workload more than the point augmentation. This suggests that a consistent tracking strategy or the development of a visually based internal model is needed to lessen workload.

Eberts, 1987 (continued)

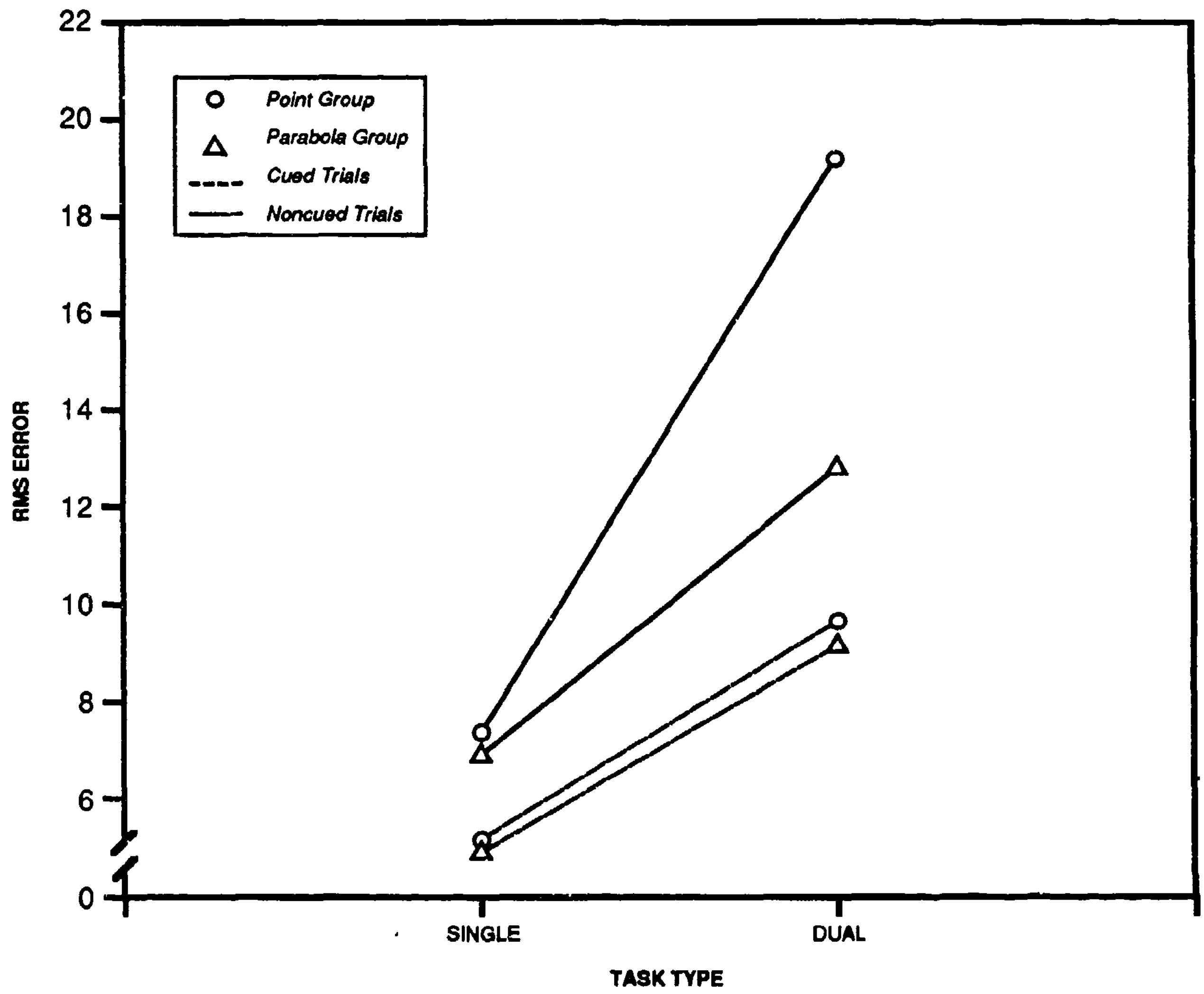


Stick velocities for the two augmented groups on the cued and noncued trials during Experiment 1.



Stick velocities for all three groups on the last eight noncued test trials of each session during Experiment 1.

Eberts, 1987 (continued)



RMS Tracking error as a function of task type (Single or dual) for the cued and noncued trials of the two augmented groups in Experiment 2.

Edwards, B.J., Weyer, D.C., & Smith, B.A. (1979). Undergraduate pilot training: Visual discrimination pre-training for landing task (AFHRL-TR-78-78). Brooks Air Force Base, Texas: Air Force Systems Command (DTIC No. AD-A068141).

Task: Visual discrimination

Taxa: Visual

Type of Data: Group

Subjects: 38 males students at Williams AFB

Design: Repeated measures

Training:

1. Subjects received either cognitive pre-training, Visual Discrimination Pre-training (VDPT), VDPT and Advance Simulator for Pilot Training (ASPT), only ASPT, or no pre-training.

Independent Variable(s):

1. Pre-training (multimedia or no pre-training)

Dependent Measure(s):

1. Instructor-rated transfer performance in the T-37 aircraft

Abstract: This study examined the utility of training task-relevant visual discrimination stalls as prerequisite behaviors for subsequently taught landing skills. A multimedia training package was developed and used to teach visual skills to student pilots prior to their training for landing procedures in the T-37 aircraft. Transfer of visual discrimination skills of the landing field environment was assessed by measuring students' landing skills in the Advanced Simulator for Pilot Training and in the T-37 aircraft. Results indicated that pre-trained and non-pre-trained performance was not significantly different.

Overall Performance Rating

by IP for Final Turn Trials in Aircraft

Group 1 Control	Group 2 Procedures Only	Group 3 VDPT	Group 4 ASPT	Group 5 VDPT/ASPT
17	19	24	14	32.5
12	8	36	2	9
34.5	10	5	30	21
3.5		3.5	15.5	19
37	22.5	1	25	22.5
19	13	15.5	6.5	11
26.5	32.5	6.5	28	34.5
	38	30	30	26.5
R^1=154.5	R^2=161.5	R^3=125.5	R^4=156.5	R^5=182
H=1.88 ($p < .75$)				

Fisk, A.D., & Lloyd, S.J. (1988). The role of stimulus-to-rule consistency in learning rapid application of spatial rules. Human Factors, 30, 35-49.

Task: Decision making

Taxa: Problem solving

Type of Data: Group

Subjects: 6 undergraduates

Design: Repeated measures

Training:

1. Subjects repeated a decision-making task over 6048 trials.

Independent Variable(s):

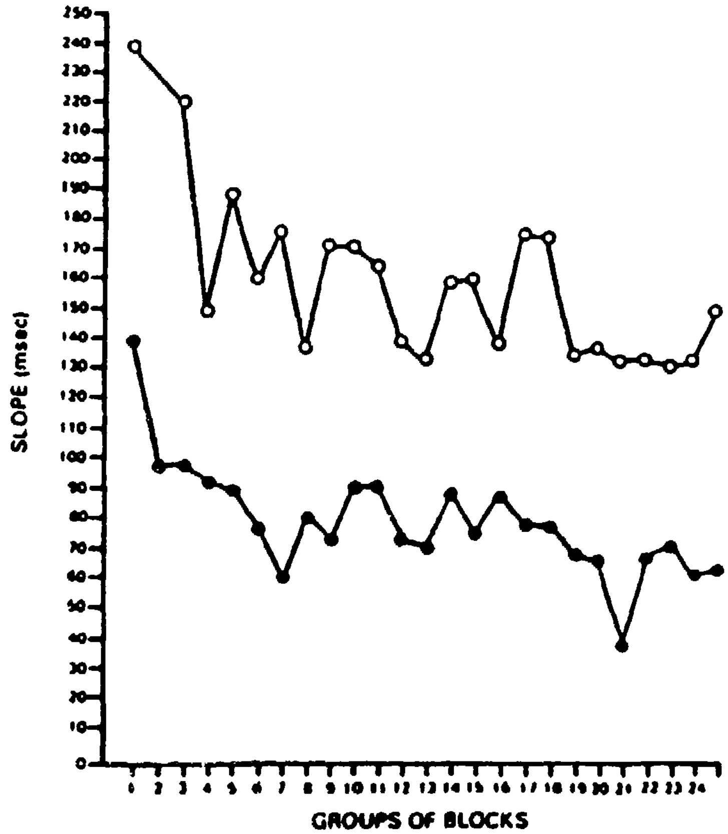
1. Memory set size

2. Possibility of correct decision (positive or negative trials)

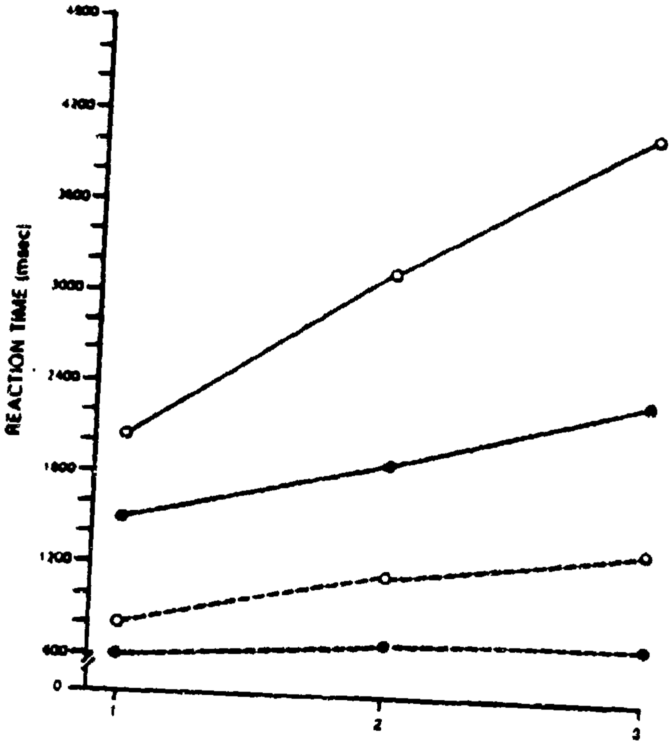
Dependent Measure(s):

1. Reaction time in dual-task situation

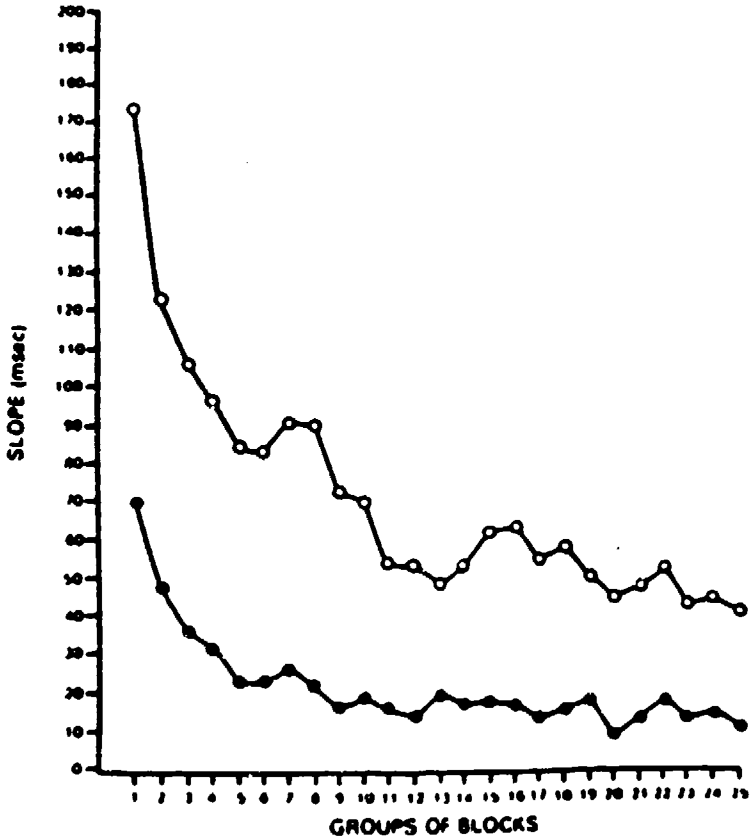
Abstract: This study addressed the effects of intercomponent consistency on skill acquisition in a class of cognitively demanding tasks that required rapid integration of information and rapid rule application. After extensive practice, subjects' performance was remarkably similar to performance observed in traditional perceptual learning tasks. This suggests that mechanisms underlying traditional learning in visual search and rule-based spatial learning are similar. Subjects who were trained so that consistent stimulus to rule association were built up and strengthened with practice performed qualitatively and quantitatively better than those who were trained with inconsistent stimulus to rule associations.



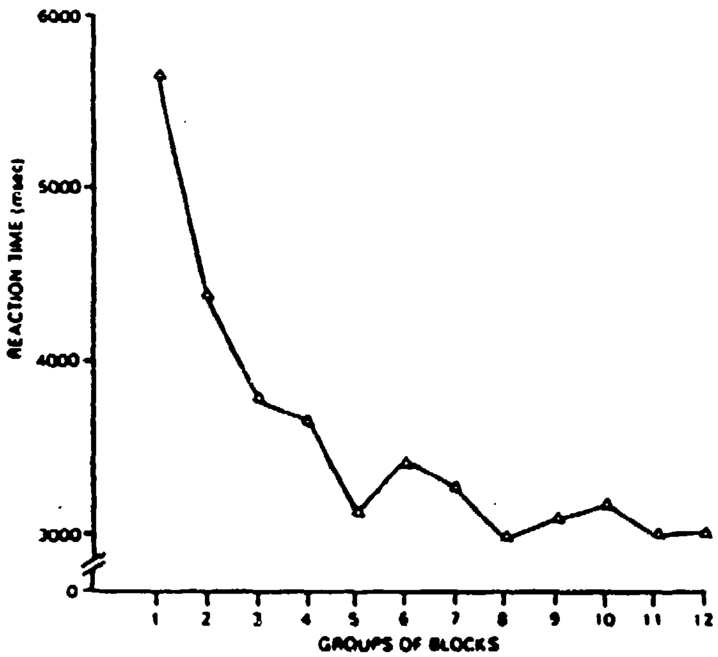
Average comparison slope estimates (in ms) for positive trials (solid circles) and negative trials (open circles). The data are shown as a function of practice (groups of blocks). Data from Experiment 3A.



. *Reaction time plotted as a function of memory-set size early in practice (solid lines) and late in practice (dashed lines). Solid circles represent average correct-trial reaction time when a capture was possible (positive trials), and the open circles show reaction time when no capture was possible (negative trials). Note that memory-set sizes 1, 2, and 3 correspond to a comparison load of 6, 12, and 18. Data from Experiment 2A.*



Average comparison slope estimates (in ms) for positive trials (solid circles) and negative trials (open circles). The data are shown as a function of practice (groups of blocks). Data from Experiment 2A.



Reaction time to indicate correctly the correct capturing "game-piece." The data are plotted as a function of practice (groups of blocks). Data from Experiment 1.

Johnson, S.L. (1978) Retention and transfer of training on a procedural task; interaction of training strategy and cognitive style (AFOSR-TR-78-1161). Bolling AFB, DC: Air Force Office of Scientific Research.

Task: Set up a conveyor line

Taxa: Information processing/Motor

Type of Data: Group

Subjects: 60 (36 male, 24 female) paid subjects

Design: Repeated measures (initial training, retraining, and transfer)

Training:

1. Subjects were trained under one of three conditions: conventional, reproduction, or blind.

Independent Variable(s):

1. Initial training strategy for "checklist" sequential task: conventional, reproduction, or blind strategy

Dependent Measure(s):

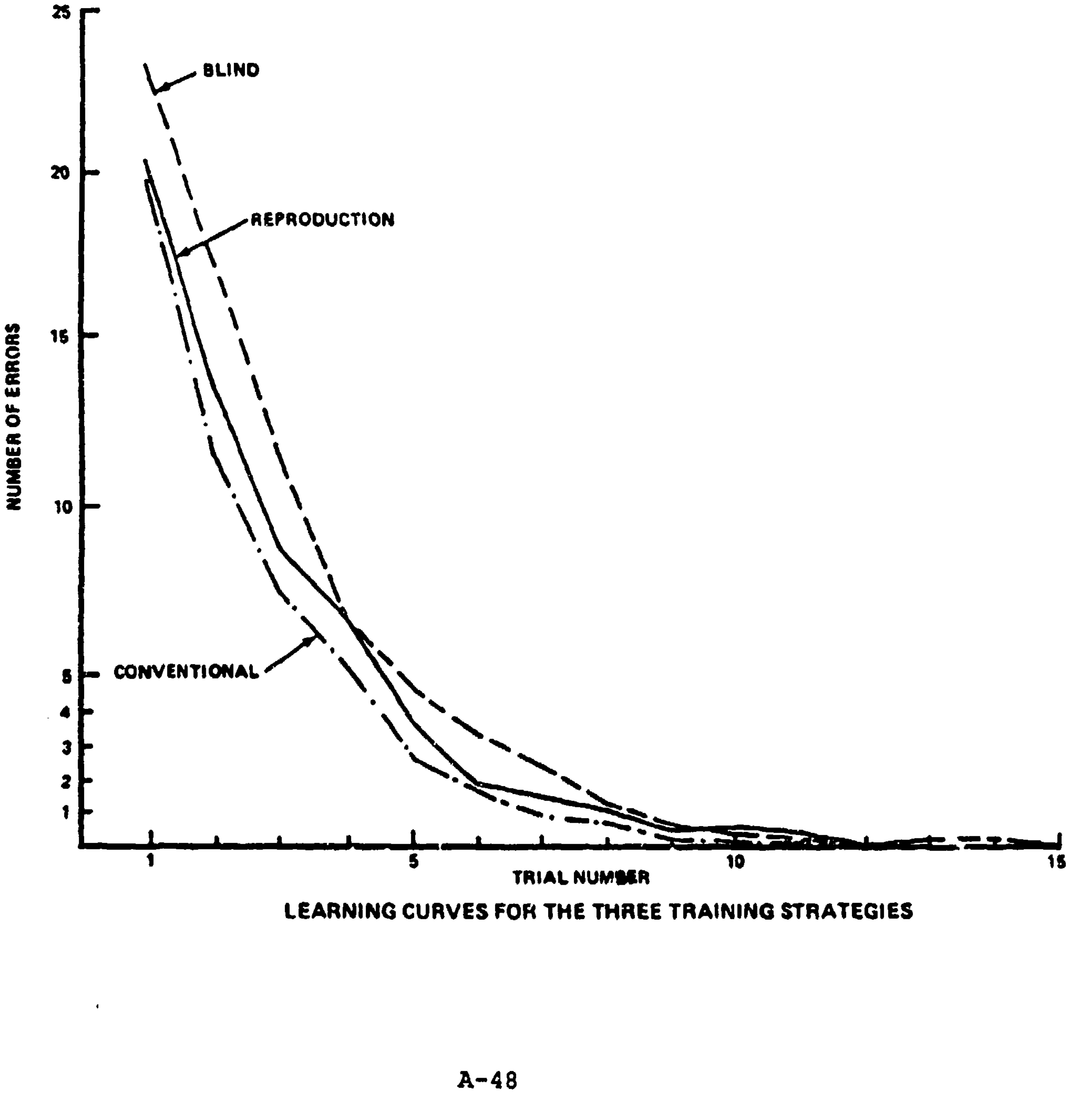
1. Sequence and setting errors on sequential task

2. Time to perform the last trial on sequential task

3. Transfer errors on actual sequential task (e.g., control panel equipment)

Abstract: This study examined the effectiveness of three different training strategies related to initial training, retention, and transfer of training. Also, the possibility of matching the instructional strategy and the trainees cognitive style was evaluated. The task used was representative of the many sequential procedures performed which range from operating master control panels in industrial plants to emergency procedures in air vehicles. Results indicated that a) vividness of imagery interacts with training strategy, b) training devices do not need high fidelity to be effective in training procedural tasks, and c) the use of training strategy that requires self-cuing and self-feedback is effective in increasing the retention of procedure-follow skills independent of cognitive style.

Johnson, S.L. (1978)



LEARNING CURVES FOR THE THREE TRAINING STRATEGIES

Kanfer, R., and Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. Journal of Applied Psychology Monograph, 74, 657-690.

Task: Air traffic controller task (ATC)

Taxa: Perceptual; Fine motor; Information processing/ Problem solving

Type of Data: Group

Subjects: 552 U.S. Air Force recruits

Design: Repeated measures

Training:

1. Part-task training on the ATC using either a procedural or declarative strategy
2. Subjects were given "goal" or "no-goal" instructions
3. Final Test - 6 Trials on the full ATC task

Independent Variable(s):

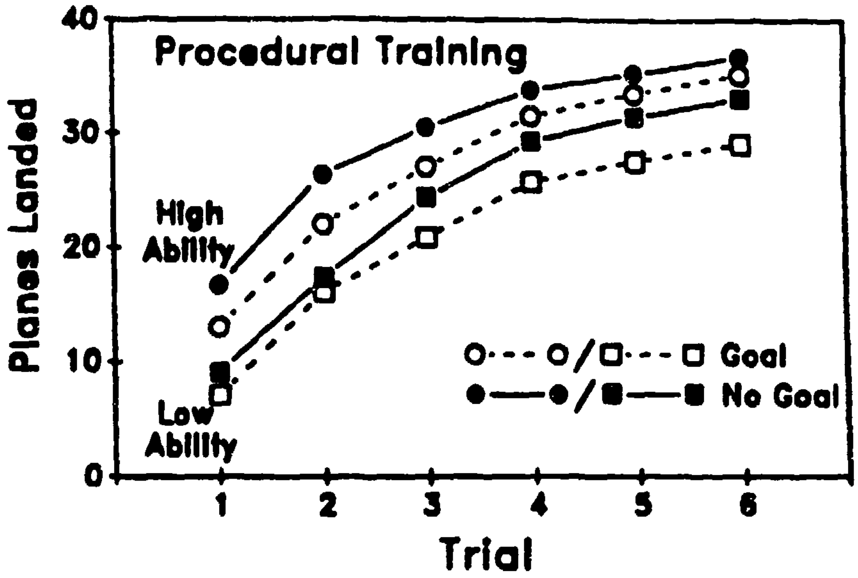
1. ATC Part-task training strategy: declarative or procedural
2. Goal setting strategy: prior to attempting the full ATC task, subjects were either given "goal" or "no-goal" instructions
3. Individual ability as measured by ASVAB test

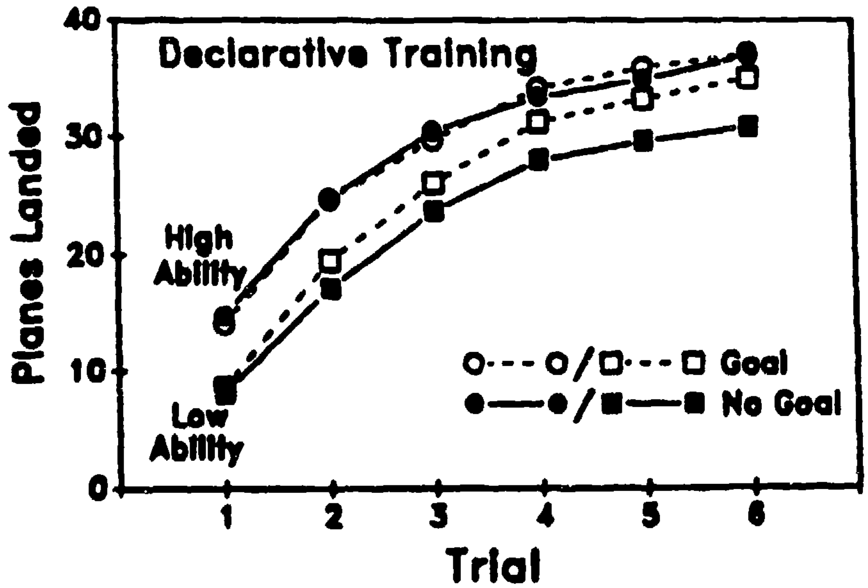
Dependent Measure(s):

1. Number of planes landed on six trials of full ATC

Abstract: This study simultaneously considered individual differences in cognitive abilities, self-regulatory processes of motivation, and information processing demands. Evidence for this framework is provided in the context of skill acquisition in which information processing and ability demands change as a function of practice, training paradigm, and timing of goal setting. Three field based experiments were conducted with 1,010 U.S. Air Force trainees. In Experiment 2, the basic ability-performance parameters of air traffic controller task and goal setting effects late in practice were examined. The results have implications for notions of ability-motivation interactions and design of training and motivational programs.

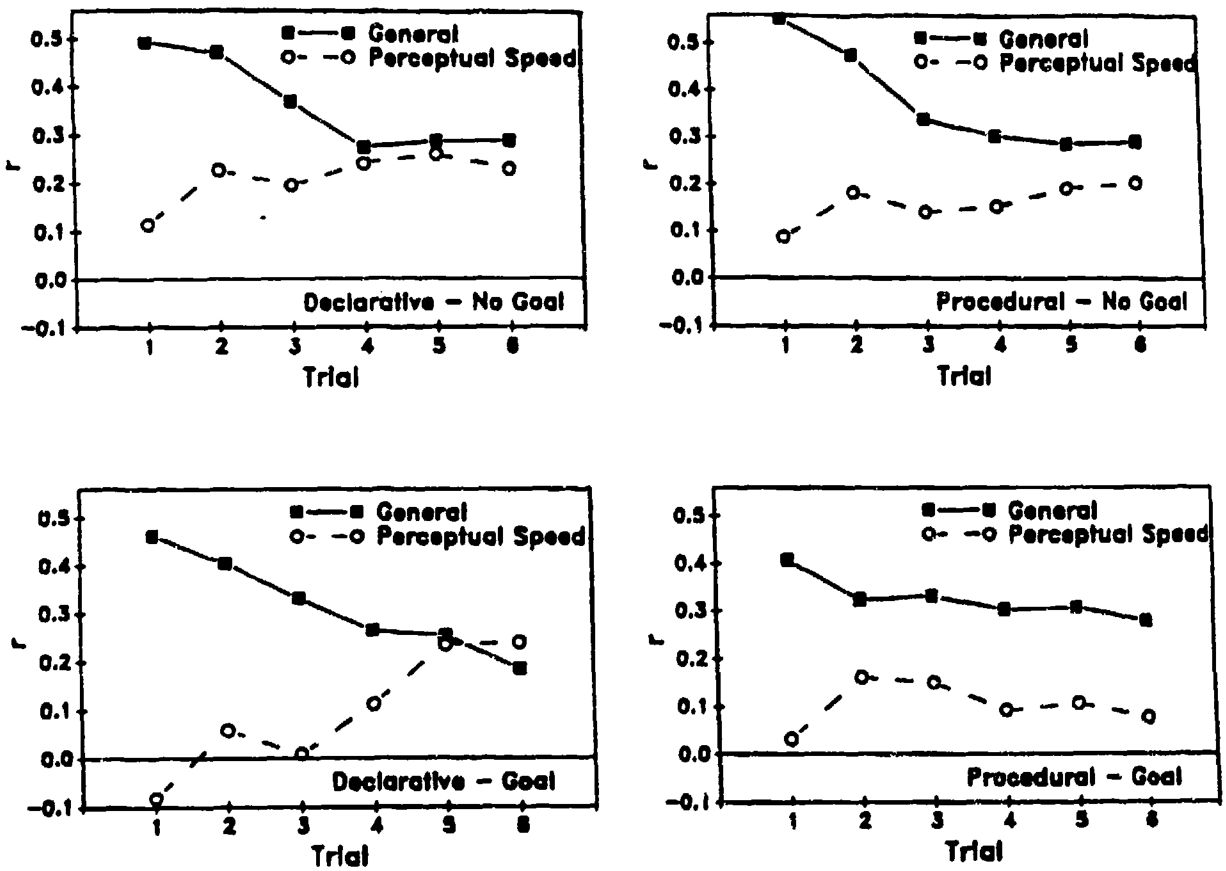
Kanfer et al. 1989 (continued)



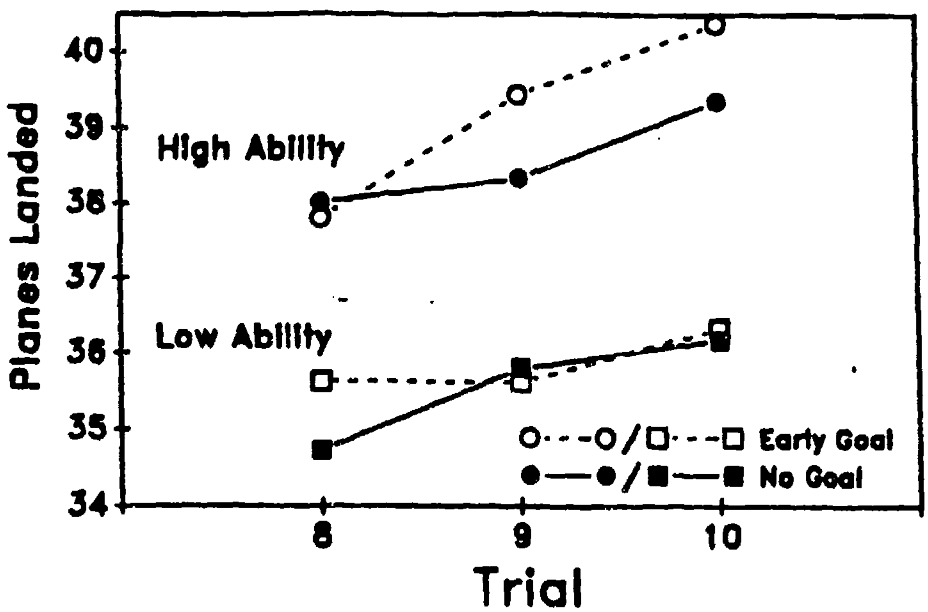


Planes landed as a function of full air traffic controller task trial, by condition and by ability group (median split). (No-goal conditions are in solid lines; goal conditions are in dotted lines. High-ability group is in circles; low-ability group is in squares.)

Kanfer et al. 1989 (continued)

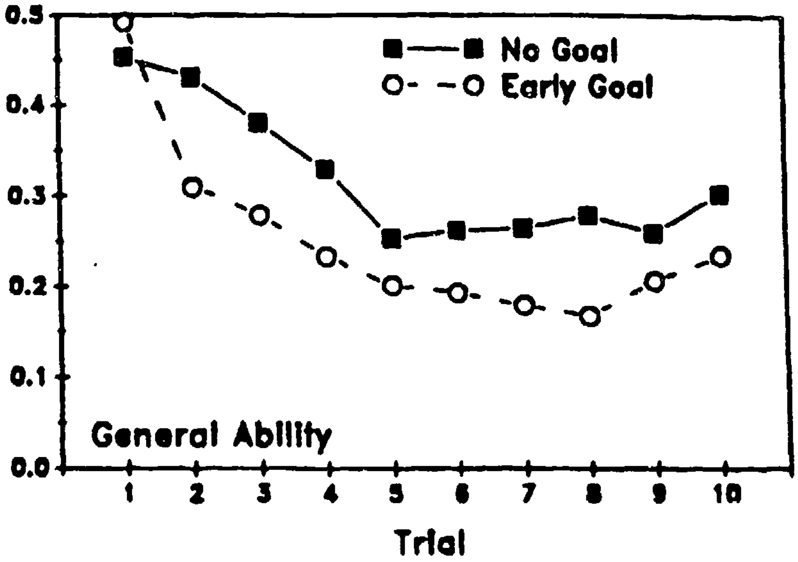


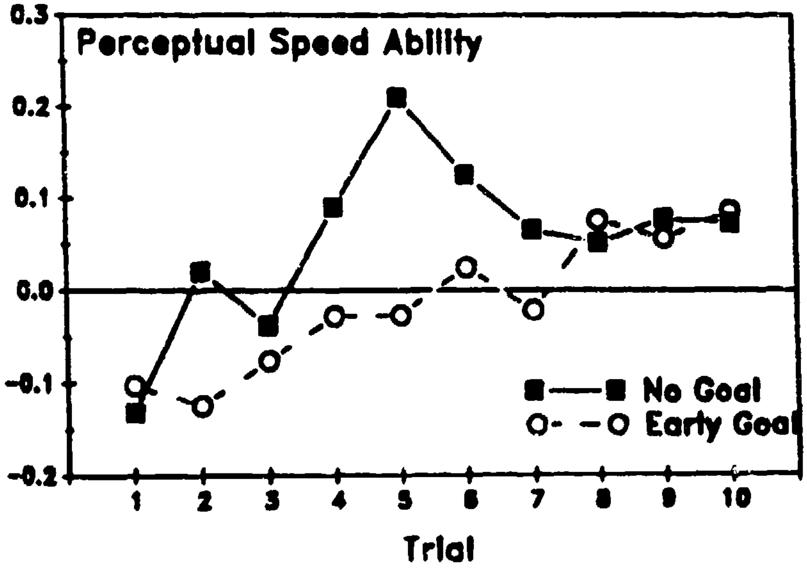
Ability-performance correlations by ability factor, condition, and full-task practice trial. (General ability-performance correlations are in solid lines and filled squares. Perceptual speed-performance correlations are in dashed lines and open circles.)



Planes landed as a function of air traffic controller task trial for final three trials, by condition and by ability group (median split). (No-goal condition is in solid lines; early-goal condition is in dashed lines. High-ability group is in circles; low-ability group is in squares.)

Kanfer et al. 1989 (continued)





Lintern, G., Roscoe, N.R., Koonce, J.M., & Segal, L.D. (1990). Transfer of landing skills in beginning flight training. Human Factors, 32, 319-327.

Task: Aircraft landing task

Taxa: Motor/Visual

Type of Data: Group

Subjects: 42 flight students with no previous flight experience

Design: Repeated measures

Training:

1. Subjects either received traditional instruction or traditional instruction and supplemental simulator training.

Independent Variable(s):

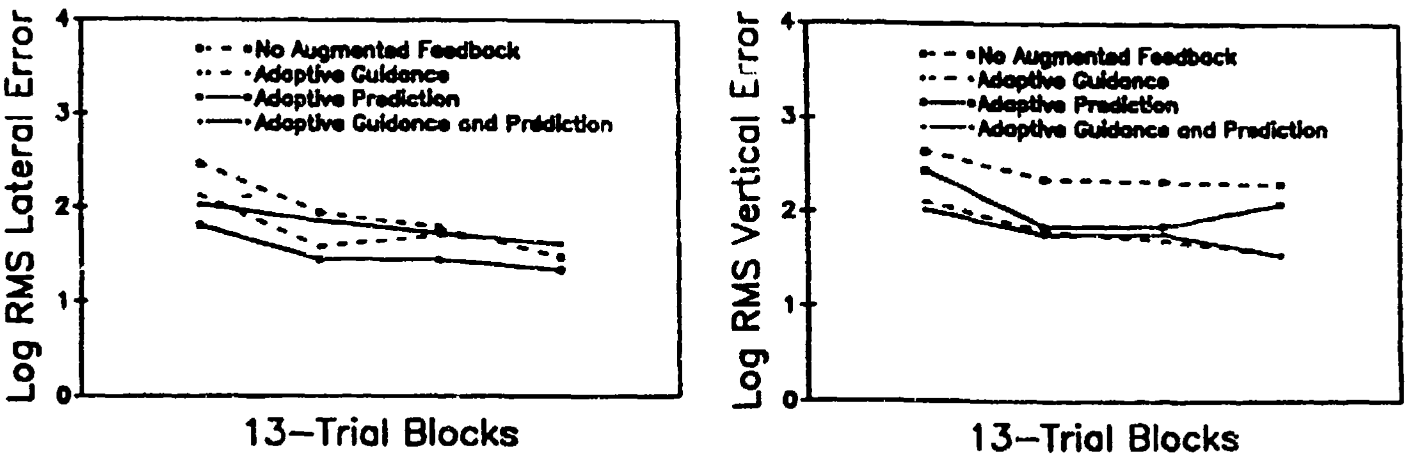
1. Landing instruction: supplemental simulator training or traditional training

Dependent Measure(s):

1. Raw tracking error on landing task

Abstract: Beginning flight students were given two sessions of landing practice in a simulator with computer animated contact landing display before commencing landing practice in the aircraft. Subjects in the experimental condition received supplemental simulator landing training while subjects were not afforded this treatment. Results demonstrated that those students in the experimental group required significantly fewer pre-solo landings in the airplane than did those in the control group. Also, some students were provided adaptive visual augmentation during simulator training which produced incremental transfer effects attributable to this feature.

Lintern et al. 1990 (continued)



Lineup (left) and glideslope (right) performance in the simulator for the initial approach to landing.

Malloy, T.E., Mitchell, C., & Gordon, O.E. (1987). Training cognitive strategies underlying intelligent problem solving. Perception and Motor Skills, 64, 1039-1046.

Task: Problem solving task

Taxa: Problem solving

Type of Data: Group

Subjects: 54 introductory psychology students

Design: Repeated measures

Training:

1. Subjects either received different degrees of problem solving strategy training or received no special treatment.

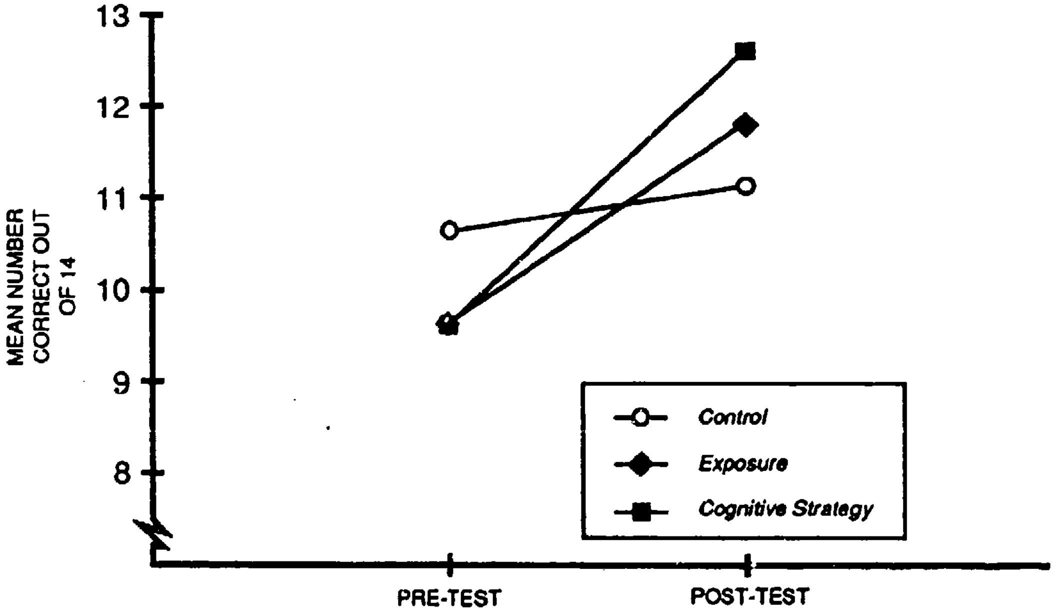
Independent Variable(s):

1. Cognitive strategy training: using either Raven's Progressive Matrices Test, brief exposure to the test, or no exposure

Dependent Measure(s):

1. Problem solving performance on 17 matrix puzzles

Abstract: This study examined the extent to which training problem solving strategies affects performance. Cognitive strategies underlying excellent performance of intelligent people on the Raven's Progressive Matrices Test were used to teach subjects in the experimental conditions. Subjects in the control group were not exposed to the training package at all. Results showed that from pretest to post-test, the group who received problem solving training performed significantly better than those who were not trained. Secondly, those who received problem solving training performed better than those who were not trained on a Piagetian task suggesting a broad improvement in cognitive functioning.



Mean number of correct solutions out of a possible 14
as a function of type of training and test

Masson, M.E. (1986). Identification of typographically transformed words: Instance-based skill acquisition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12, 479-488.

Task: Word search

Taxa: Visual/problem solving

Type of Data: Group

Subjects: 24 university students

Design: Repeated measures

Training:

1. Students were presented a randomly ordered list of 24 word triplets and were then asked to identify individual words. A test phase ensued immediately after training and a similar procedure was repeated.

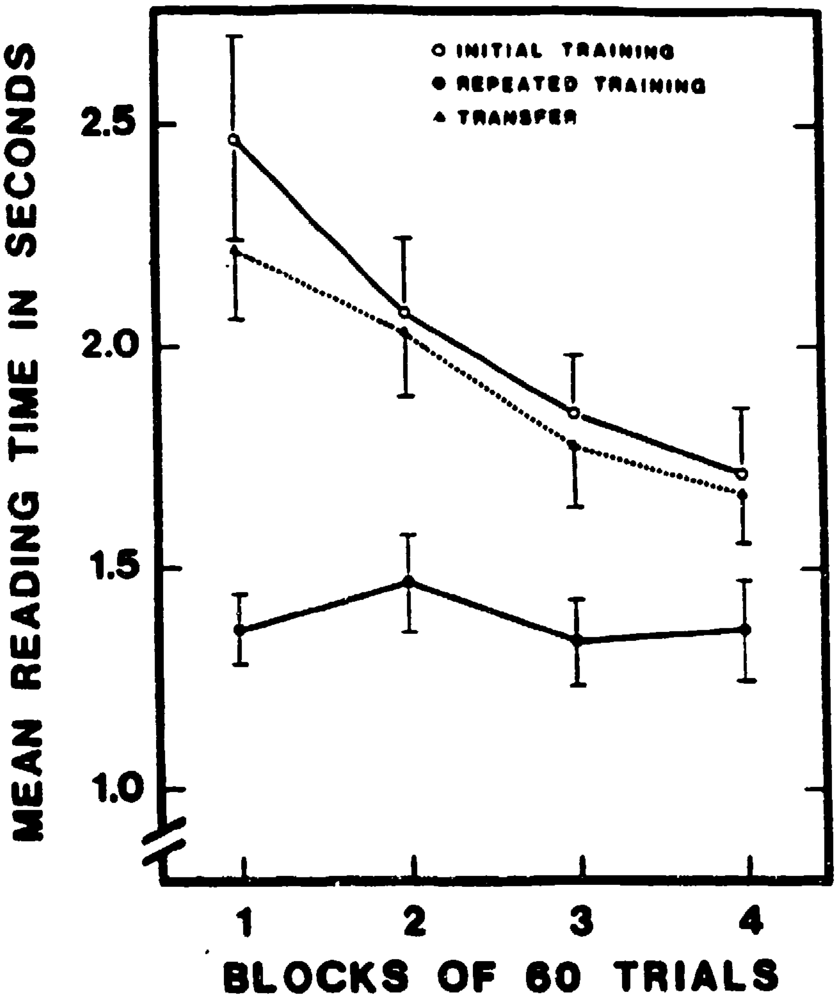
Independent Variable(s):

1. Extent of exposure to word triplets

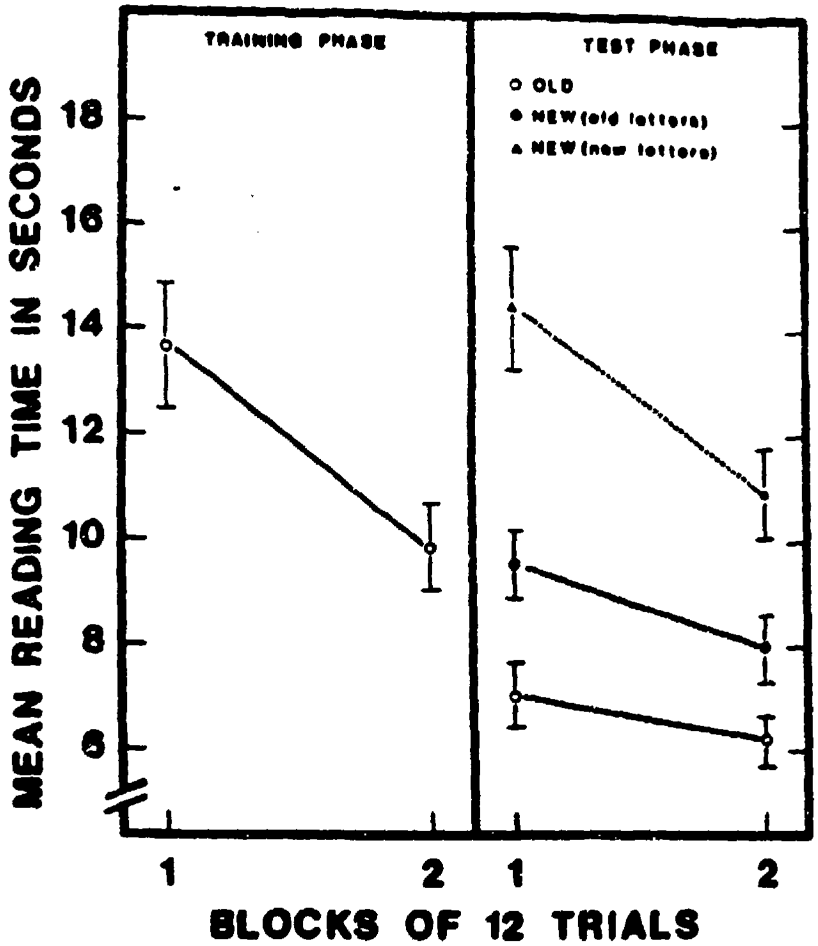
Dependent Measure(s):

1. Word identification score

Abstract: This study demonstrated that the transfer of word identification skill was highly specific and occurs only when training and test instances shared common letters printed in the same case. Also, transfer of skill depended on the visual patterns formed by adjacent letters and word shape. Presentation of a word in training and test phases significantly improved test phase identification even if a unique visual pattern was used. The results are consistent with an instance-based view of word identification skill.



Mean reading times for the training and test phases of Experiment 4. (Vertical bars indicate one standard error of the mean.)



. Mean reading times for the training and test phases of Experiment 1. (Vertical bars indicate one standard error of the mean.)

Miller, E.E. (1975). Instructional strategies using low-cost simulation for electronic maintenance (HumRRO Technical Report HumRRo-FR-WD(TX)-75-20). Alexandria, VA: Human Resource Research Organization.

Task: Electronic maintenance and problem solving on a complex system (Improved HAWK)

Taxa: Problem solving

Type of Data: Group

Subjects: 78 students enrolled in the Improved HAWK Firing Section Mechanics Program of Instruction

Design: Between subjects, no repeated measures

Training:

1. Subjects either received traditional instruction or traditional instruction and supplemental training.

Independent Variable(s):

1. Instructional training for Improved Hawk

Dependent Measure(s):

1. Practical quiz test score for Improved Hawk

2. Score on standardized quiz for Improved Hawk

Abstract: This study examined a variety of instructional strategies that employ visual aids to improve electronic maintenance training on a complex system (Improved Hawk). Large color photos of control/indicator panels, supplemented by color slides, were developed specifically for use in the classroom. Supplementary exercises were developed to be used by students while awaiting their participation on the equipment. Results indicated that subjects receiving experimental training performed significantly better with respect to checks, adjustments, and fault isolation compared with the control group.

PERFORMANCE ON FAULT ISOLATION PROBLEMS IN SECOND EXAMINATION PERIOD

Problem 1 (Crystal Detector)

	DID NOT COMPLETE IN TIME	COMPLETED IN TIME	ERRORS		
			0	1	2 or more
Experimental	3	35	7	12	19
Control	11	28	5	5	29

$\chi^2{}_1 = 5.3, p < .05$ $\chi^2{}_2 = 5.3$, n.s.

Miller, 1975 (continued)

PERFORMANCE ON FAULT ISOLATION PROBLEMS
IN FIRST EXAMINATION PERIOD

Problem 1 (Interlock)

	DID NOT COMPLETE IN TIME	COMPLETED IN TIME	ERRORS		
			0	1	2 or more
Experimental	5	34	27	10	2
Control	19	20	11	14	14

$\chi^2_1 = 13.2$, $p < .001$ $\chi^2_2 = 16.4$, $p < .001$

Problem 2 (+100 VDC Power Supply)

	DID NOT COMPLETE IN TIME	COMPLETED IN TIME	ERRORS		
			0	1	2 or more
Experimental	13	26	15	20	4
Control	28	11	6	13	20

$\chi^2_1 = 10.6$, $p < .001$ $\chi^2_2 = 15.9$, $p < .001$

Morris, N.M., & Rouse, W.B. (1985). The effect of type of knowledge upon human problem solving in a process control task. IEEE, 15, 698707.

Task: Process control task and problem solving

Taxa: Problem solving

Type of Data: Group

Subjects: 32 engineering students

Design: 2(Principles) x 2(Procedures) factorial design with Repeated measures

Training:

1. Subjects either received minimal instructions, minimal instructions plus principles, minimal instructions plus procedures, or all instructions (minimal instructions, principles, and procedures) before being tested in a control task problem solving task.

Independent Variable(s):

1. Instructional training for dynamic production process using simulator (PLANT)

Dependent Measure(s):

1. Production as measured on dynamic production process simulator (PLANT)

Abstract: This study examined the question of what the operator of a dynamic system needs to know. The study used PLANT, a simulation of a generic dynamic production process. Knowledge of PLANT was manipulated by providing different types of instructions to four different groups. There was no significant difference between groups receiving different instructions. However, subjects who received better instructions controlled the system in a more stable fashion.

CORRELATIONS BETWEEN DEPENDENT MEASURES[1]

	PROD	TRIPS	NOPEN	VAR	FIX	TEST2	SECT1	SECT2
TRIPS	−0.437[2]							
NOPEN	0.673[2]	−0.706[2]						
VAR	−0.574[2]	0.967[2]	−0.768[2]					
FIX	−0.429[2]	0.141	−0.234	0.218				
TEST2	0.19[illegible]	−0.200	0.313	−0.258	0.107			
SECT1	−0.021	0.261	−0.189	0.268	−0.100	0.148		
SECT2	0.190	−0.238	0.366[2]	−0.292	0.225	0.860[2]	0.040	
SECT3	0.105	−0.161	0.157	−0.197	−0.056	0.661[2]	−0.022	0.225

[1] PROD — average production/iteration
TRIPS — number of automatic valve trips/iteration
NOPEN — average number of valves open/iteration
VAR — variance of tank heights in PLANT
FIX — average time to diagnose valve and pump failures
TEST2 — overall score on TEST 2
SECT1, SECT2, SECT3 — scores (percent correct) on subsections of TEST 2
SECT1 — minimal questions
SECT2 — procedural questions
SECT3 — principles questions.

[2] $p < 0.05$

Myers, G.L., & Fisk, A.D. (1987). Training consistent task components: Application of automatic and controlled processing theory to industrial task training. Human Factors, 29, 255-268.

Task: Telecommunication processing task

Taxa: Information processing/Visual

Type of Data: Group

Subjects: 8 AT&T Communications employees (five males, three females)

Design: simple measurement (independent variable was only statistically manipulated)
Training:

1. Subjects were trained under either the consistent mapping or variable mapping condition.

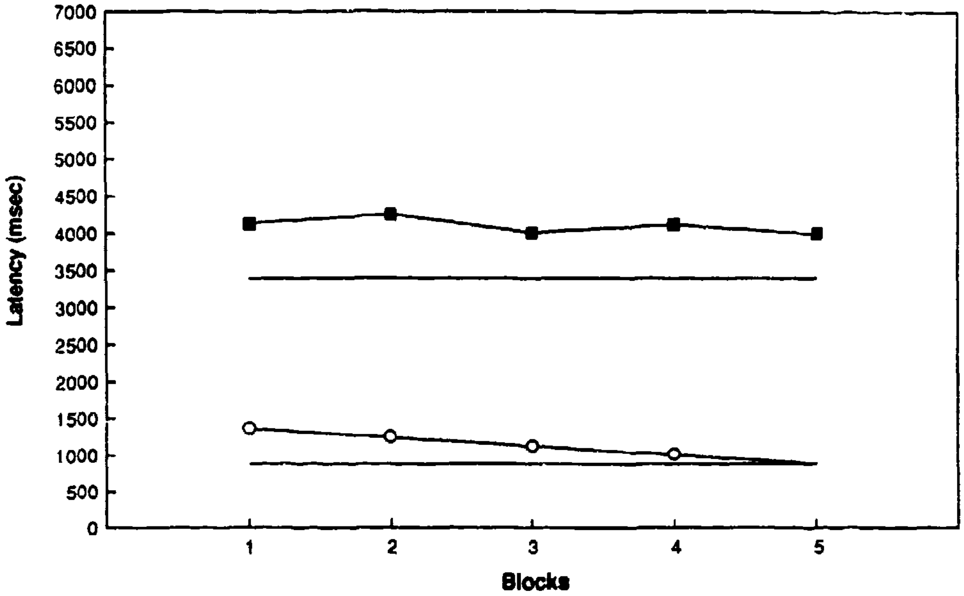
Independent Variable(s):

1. Mapping of information (variably or consistently mapped) in training for a telecommunications type of visual search task

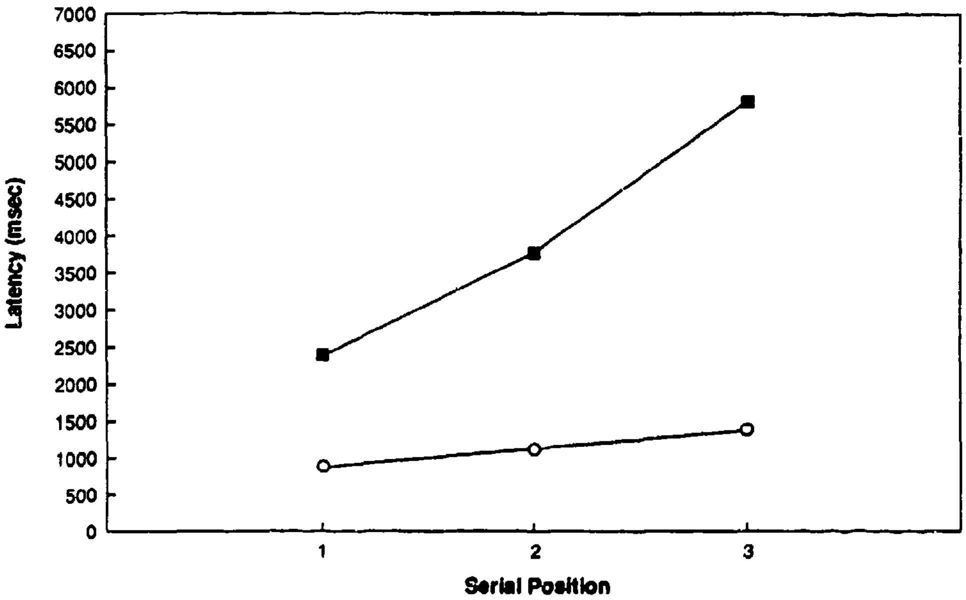
Dependent Measure(s):

1. Speed and Accuracy on telecommunications visual search task

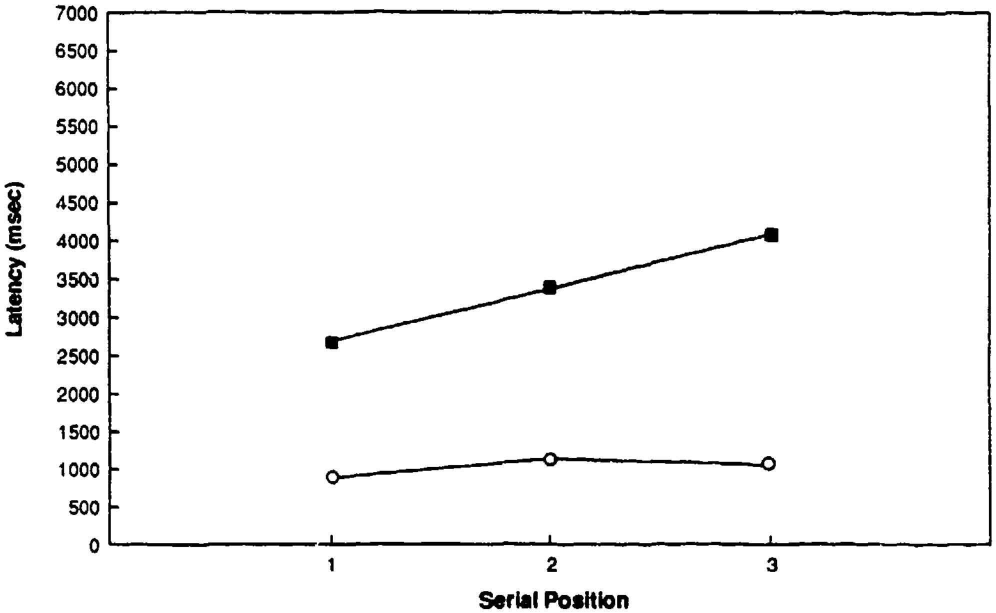
Abstract: This study examined the generality of automatic/controlled processing training principles to rich, complex tasks. Subjects' tasks were modelled after a job function performed in the telecommunications industry. The tasks required subjects to process conjunctions of information. Large quantitative and qualitative differences were reported between consistently and variably mapped training conditions. Results indicate that theories of automatic and controlled processing may be expanded to include domains of rich, complex industrial tasks.



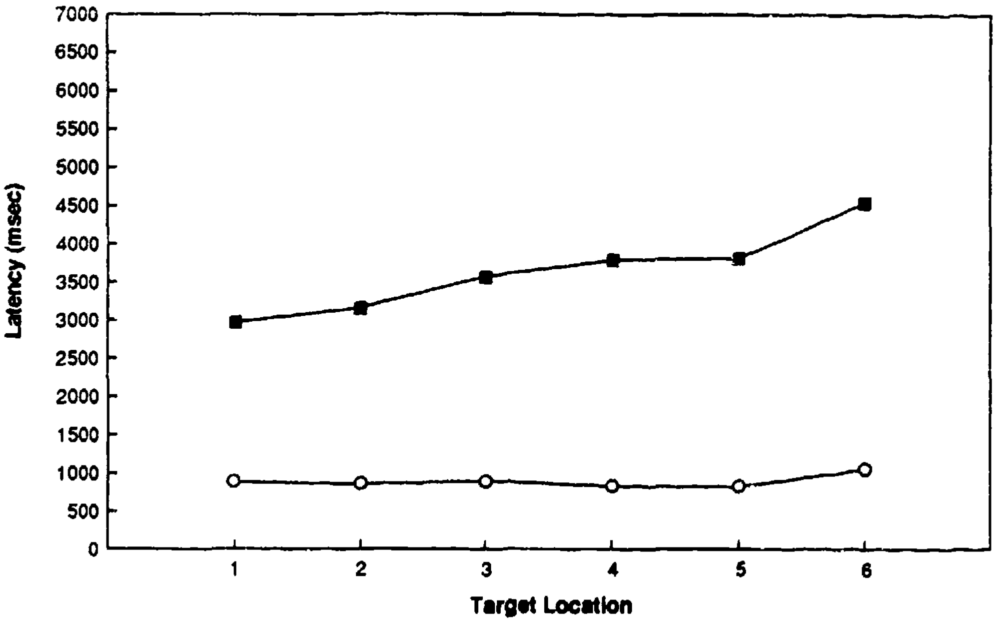
Response latency as a function of practice in Experiment 2. The solid lines with no circles or squares represent previous VM or CM performance levels. CM (open circles) = consistent attending; VM (filled squares) = inconsistent attending.



Response latency as a function of serial position for both CM (open circles) and VM (filled squares) in Experiment 2. (CM = consistent attending; VM = inconsistent attending.)



Response latency as a function of serial position for both CM (open circles) and VM (filled squares). The data are from the last group of blocks.



Response latency as a function of the target's location on the display for both CM (open circles) and VM (filled squares). Target Location 1 corresponds to the upper-left-most display position. Target Location 6 represents the lower-right-most display position. The data are from the last group of blocks

Patrick, J., and Haines, B. (1988). Training and transfer of fault-finding skill. Ergonomics, 31(2), 193-210.

Task: Fault finding task

Taxa: Problem solving/information processing

Type of Data: Group

Subjects: 24 University students

Design: Repeated measures

Training:

1. Technical story or set of diagnostic heuristics

2. Final Test - "Transfer" fault finding test using a different scenario

Independent Variable(s):

1. Training materials (technical story vs heuristics)

Dependent Measure(s):

1. Accuracy and Speed of fault diagnosis test with different scenario on transfer test scenario

Abstract: This study investigated the effect of different training materials on training and transfer of fault-finding skill. Two simulated chemical plants were used as the fault-finding domain. The training materials consisted of either a technical story or diagnostic heuristics. Both conditions improved at fault-finding during training. However, the technical story condition had better overall performance during training. Transfer was high and positive for both conditions. Results are discussed with respect to the role of theory in training fault-finding.

Accuracy and speed (mins) of diagnoses during fault-finding tests for the experimental conditions.

		Fault-finding tests			
		Test 1	Test 2	Test 3 'Old' faults	Test 3 'New (same category)' faults
Condition T(FP2)					
Accuracy	mean	0·10	0·20	0·53	0·43
	s.d.	0·11	0·19	0·30	0·20
Speed	mean	3·99	3·09	2·22	2·87
	s.d.	0·85	0·77	0·67	0·93
Condition H(FP2)					
Accuracy	mean	0·11	0·24	0·53	0·23
	s.d.	0·13	0·18	0·21	0·17
Speed	mean	4·01	2·70	2·05	2·76
	s.d.	0·83	0·89	0·52	0·55
Condition H(CD1)					
Accuracy	mean	0·22	0·22	0·38	0·30
	s.d.	0·18	0·21	0·27	0·19
Speed	mean	4·19	2·96	2·57	2·97
	s.d.	0·49	0·67	0·58	0·41

Transfer of training from FP2 to CD1 plant in terms of accuracy, speed and number o instrument readings.

		Training/test conditions		
		T(FP2)–Test 4	H(FP2)–Test 4	H(CD1)–Test 1 (control)
Accuracy				
	mean	0·68	0·60	0·22
	s.d.	0·20	0·22	0·18
Speed (min)				
	mean	2·13	2·23	4·19
	s.d.	0·51	0·72	0·49
Instrument readings				
	mean	6·08	5·88	12·47
	s.d.	1·51	1·89	6·08

Patrick et al., 1988

Accuracy and speed (mins) of diagnoses during fault-finding tests for the experimental conditions.

		Fault-finding tests			
		Test 1	Test 2	Test 3 'Old' faults	Test 3 'New (same category)' faults
Condition T(FP2)					
Accuracy	mean	0·10	0·20	0·53	0·43
	s.d.	0·11	0·19	0·30	0·20
Speed	mean	3·99	3·09	2·22	2·87
	s.d.	0·85	0·77	0·67	0·93
Condition H(FP2)					
Accuracy	mean	0·11	0·24	0·53	0·23
	s.d.	0·13	0·18	0·21	0·17
Speed	mean	4·01	2·70	2·05	2·76
	s.d.	0·83	0·89	0·52	0·55
Condition H(CD1)					
Accuracy	mean	0·22	0·22	0·38	0·30
	s.d.	0·18	0·21	0·27	0·19
Speed	mean	4·19	2·96	2·57	2·97
	s.d.	0·49	0·67	0·58	0·41

Table 2. Diagnostic heuristics for FP2 plant.

1. Location—find the general area of failure by locating the disturbance furthest upstream.
2. Check for a failure in one of the pumps which would produce a rapid change in flow.
3. Check all control loops in the disturbed area to ensure that valves are in the appropriate positions.
4. If the failure is in the reactor/heat exchange area determine whether it is either in the reactor or the heat exchange system. A failure in the heat exchange system will produce symptoms in column A but *not* in column B.
5. If the failure is in the feed system, check whether it is in stream X or Y. If *both* feeds are disturbed, the fault is in feed Y. If *only* feed X is disturbed, the fault is in feed X.
6. Check if duplicate instruments provide the same readings. If they do not, there is an instrument reading fault.

Design of experiment.

	Experimental conditions		
	T(FP2)	H(FP2)	H(CD1)
Training			
1. Module 1: pretraining	FP2	FP2	CD1
2. Module 2: plant overview (training)	FP2	FP2	CD1
3. Fault-finding Test 1 (8 faults)	FP2	FP2	CD1
4. Module 3: training materials	Technical story for FP2	Diagnostic heuristics for FP2	Diagnostic heuristics for CD1
5. Fault-finding Test 2 (8 faults)	FP2	FP2	CD1
6. Training in fault-finding	FP2	FP2	CD1
7. Fault-finding Test 3 (12 faults—5 'old', 5 'new (same category)' and 2 'new (different category)')	FP2	FP2	CD1
Transfer			
8. Module 4: plant overview (transfer)	CD1	CD1	FP2
9. Fault-finding Test 4 (8 faults)	CD1	CD1	FP2

Rabbit, P., Cumming, G., and Vyas, S. (1979). Improvement, Learning and Retention of Skill at Visual Search. <u>The Experimental Psychology Society</u>, <u>31</u>, 441-460.

Task: Visual search

Taxa: Perception/visual discrimination

Type of Data: Individual

Subjects: 3 Experimental psychologists

Design: Repeated measures

Training:

1. 28 Practice trials on visual search task

2. Final Test - "Transfer" task consisting of discriminating targets from different background

Independent Variable(s):

1. Length of practice on task requiring visual search for embedded letters

Dependent Measure(s):

1. Reaction time for visual search of target letters

Abstract: This experiment examined factors responsible for improvement in visual search. Specifically, the study investigated the visual performance of very highly practiced subjects. Results demonstrated that after 25 days of training, variations in the size of the target set no longer affected visual search time. This may be explained in terms of learning of specific cue systems since subjects showed perfect transfer to displays on which they searched for the same target items among new background items.

Rabbit et al., 1979 (continued)

Reaction times in ms for different sub-sets of data from forefinger and second finger

	2-Targets	4-Targets	6-Targets
Forefingers			
Forefingers repeated	352	356	360
One forefinger after the other	386	389	394
Second fingers			
Second finger repeated		369	373
Second finger after other second finger		390	398
Second finger after forefinger of same hand		385	397
Second finger after forefinger of different hand		422	431
Second finger after third finger of same hand		414	427
Second finger after third finger of different hand		435	446

Rabbit et al., 1979 (continued)

Reaction times and sigma in ms for last 5 days of practice; 6-target 8-letter display condition and after transfer

	Day 26	Day 27	Day 28	Day 29	Day 30	Day 31 transfer
Mean RT	604	532	598	554	559	562
Sigma	82	54	60	77	53	58

Reaction times and sigmas in ms for different fingers

		Forefinger	Second finger	Third finger
2-Target condition	Mean RT	369		
	Sigma	52		
4-Target condition	Mean RT	386	402	
	Sigma	54	61	
8-Target condition	Mean RT	395	423	426
	Sigma	56	60	62

Reaction times and sigmas in ms in each of nine conditions

No. of target symbols sought	No. of background symbols on display	Days 1-5 RT means & sigma	Days 6-10 RT means & sigma	Days 11-15 RT means & sigma	Days 16-20 RT means & sigma	Days 21-25 RT means & sigma	Days 26-30 RT means & sigma
2	3	632	515	486	424	404	369
		σ 131	σ 111	σ 98	σ 98	σ 81	σ 52
	5	756	702	532	504	471	441
		σ 149	σ 123	σ 112	σ 92	σ 90	σ 56
	8	1010	878	699·75	606	542	537
		σ 198	σ 134	σ 122	σ 99	σ 91	σ 81
4	3	873	620	511	453	426	394
		σ 212	σ 151	σ 99	σ 104	σ 66	σ 51
	5	1021	712	582	497	451	439
		σ 209	σ 152	σ 104	σ 100	σ 52	σ 52
	8	1240	935	648	564	507	499
		σ 224	σ 124	σ 114	σ 78	σ 64	σ 54
6	3	914	707	673	520	487	426
		σ 218	σ 184	σ 104	σ 98	σ 72	σ 60
	5	1061	867	694	602	584	501
		σ 251	σ 192	σ 114	σ 98	σ 74	σ 69
	8	1270	997	776	618	593	569
		σ 242	σ 198	σ 151	σ 72	σ 65	σ 66

Reaction times and sigmas in ms for first and last runs

	No. of symbols on display	First Runs: No. of targets for which search was conducted: 2	4	9	No. of symbols on display	Last Runs: No. of targets for which search was conducted: 2	4	9
Subject	2	704	741	918	2	508	529	559
G.D.C.		98	98	118		48	54	41
	4	811	894	1123	4	540	538	557
		115	111	128		58	56	54
	9	1108	1285	1412	9	614	569	591
		142	156	159		62	60	64
Subject	2	763	842	987	2	373	473	450
S.M.V.		108	104	118		47	50	50
	4	887	1008	1113	4	497	445	497
		117	112	180		52	49	48
	9	1180	1367	1603	9	599	672	774
		156	171	214		55	52	51
Subject	2	631	796	894	2	461	447	401
B.R.		62	111	124		36	37	29
	4	741	962	1121	4	463	467	401
		117	124	137		36	49	32
	9	1026	1280	1466	9	541	626	540
		136	149	154		38	47	42

Reaction times and sigmas in ms and percentage of errors made in Experiment I

			First practice session	Last practice session	Recall session (Re-test or transfer)
2 weeks delay between practice and recall	Retest group (10 subjects)	Mean RT	1108	896	899
			(σ 149)	(σ 131)	(σ 152)
		% Errors	2·4%	1·1%	1·8%
	Transfer group (10 subjects)	Mean RT	1086	909	946
			(σ 154)	(σ 114)	(σ 144)
		% Errors	1·9%	2·1%	1·9%
4 weeks delay between practice and recall	Retest group (10 subjects)	Mean RT	1094	885	807
			(σ 162)	(σ 128)	(σ 137)
		% Errors	1·6%	2·4%	2·7%
	Transfer group (10 subjects)	Mean RT	1114	882	956
			(σ 186)	(σ 132)	(σ 159)
		% Errors	3·1%	1·8%	1·9%
6 weeks delay between practice and recall	Retest group (10 subjects)	Mean RT	1081	914	948
			(σ 134)	(σ 148)	(σ 169)
		% Errors	1·2%	2·1%	3·1%
	Transfer group (10 subjects)	Mean RT	1104	894	971
			(σ 142)	(σ 156)	(σ 182)
		% Errors	1·9%	1·9%	2·8%

Roberts, T.L., and Moran, T.P. (1983). The evaluation of text editors: Methodology and empirical results. <u>Communications of the ACM</u>, <u>26</u>, 265-283.

Task: Edit text

Taxa: Motor; Information processing/Problem solving

Subjects: 4 Novice computer users

Training:

1. Each subject trained on 9 different text editors

Independent Variable(s):

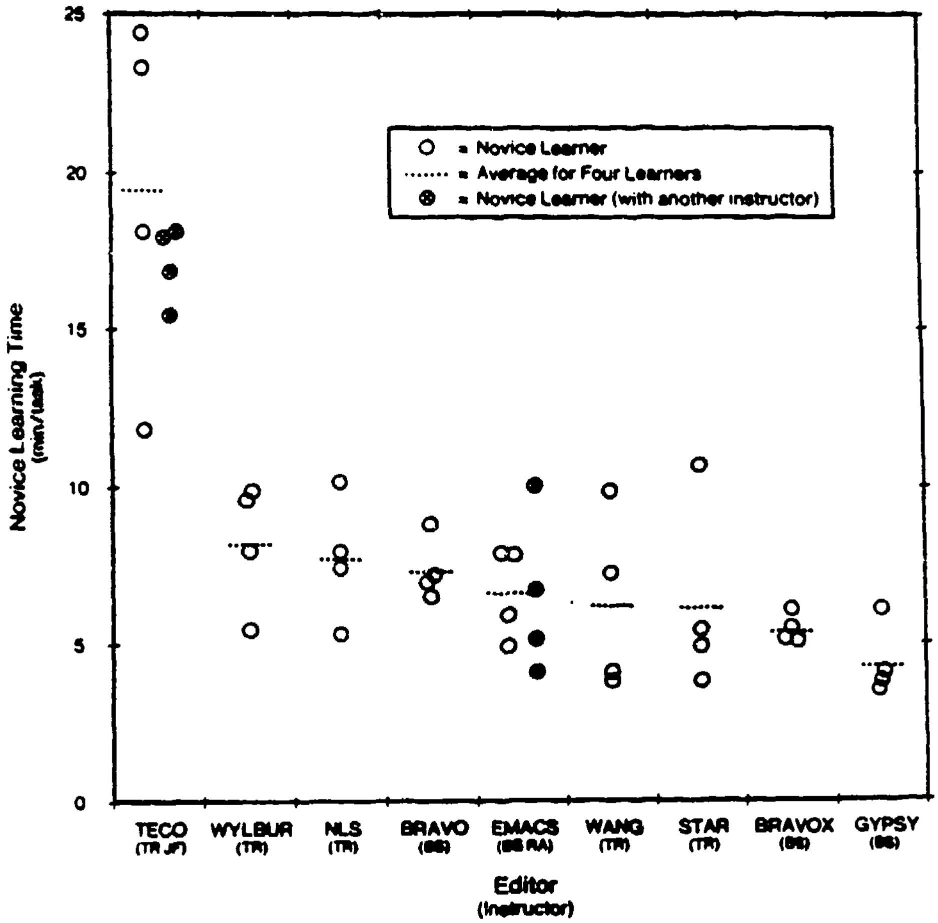
1. Experience with text editors: Novice vs Expert

Dependent Measure(s):

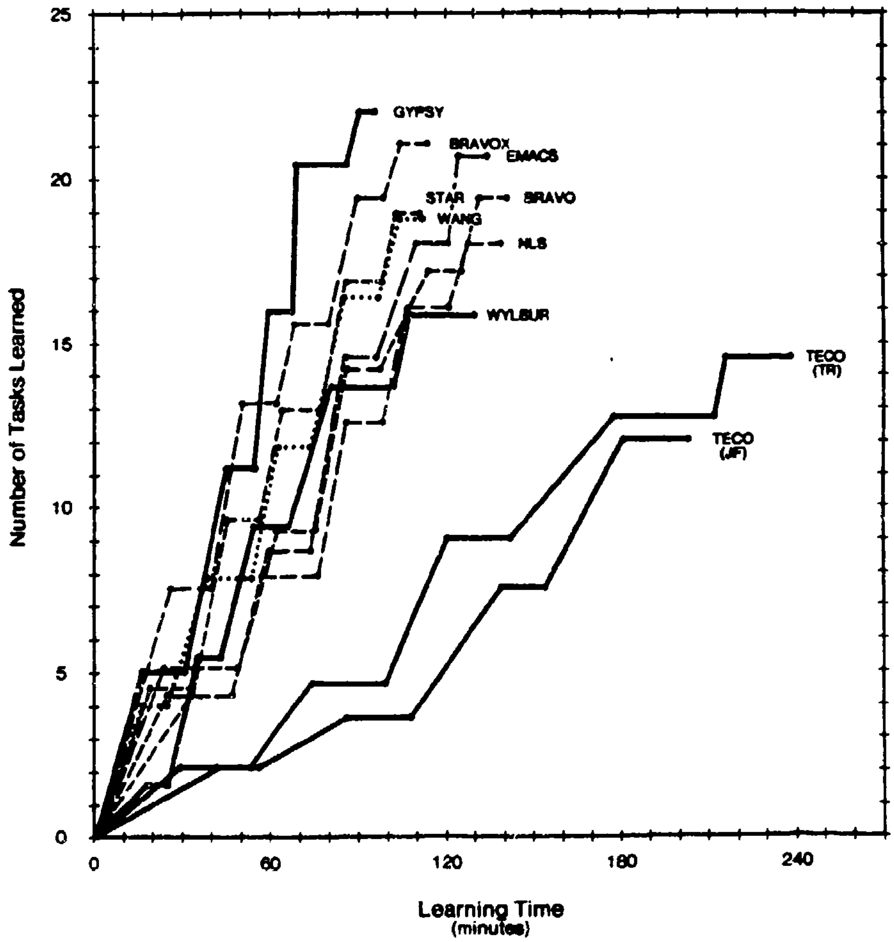
1. Time-score for each text editor
2. Individual error score for each text editor

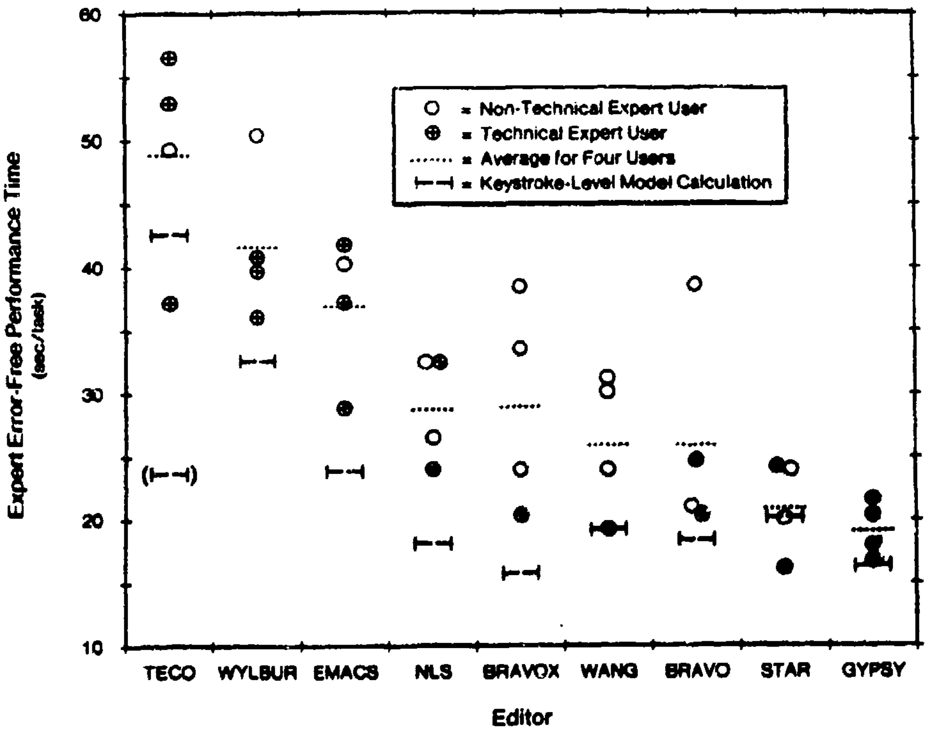
Abstract: This study presents a methodology for evaluating text editors on the following dimensions: time required for experts to perform basic editing tasks, time used by experts to make and correct errors, rate of novice learning of basic text editing tasks, and functionality of editors over more complex tasks. Time, errors, and learning are measured experimentally; functionality is measured analytically. Nine diverse editors were evaluated producing an initial database of performance results. The database not only provides information pertaining to the editors, but also provides useful information about the users - the magnitude of individual differences and factors affecting novice learning.

Learning Scores for Individual Novice Learners. The editors are ordered by descending Learning score. The instructors are noted below each editor.

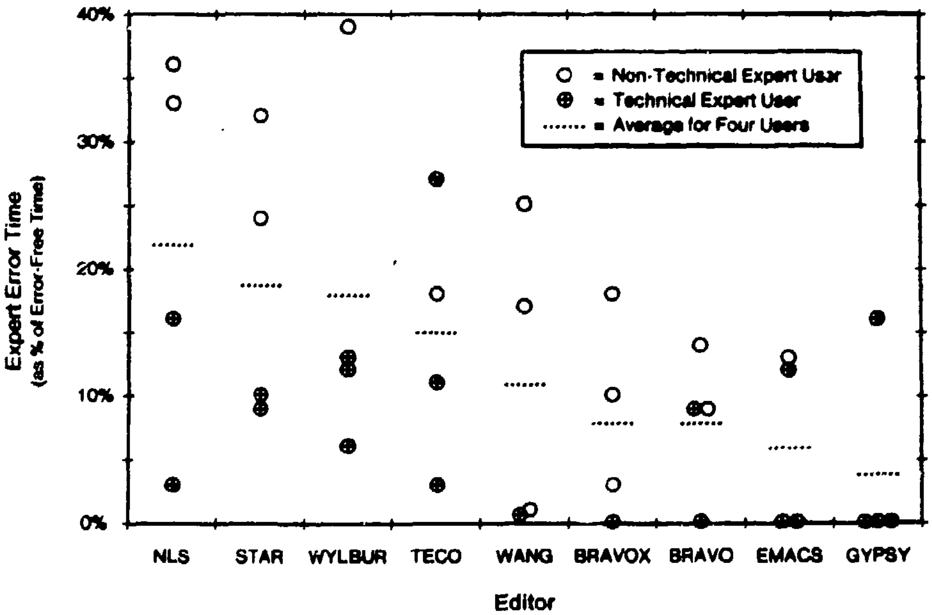


Average Learning Curves over all Learners on each Editor. The two TECO curves were produced by different inutructors.





Error-Free Time Scores for Individual Expert Users. The editors are ordered by descending Time score.



Error Time Scores for Individual Expert Users. The editors are ordered by descending Error score.

Rouse, W. B. (1978). Human problem solving performance in a fault diagnosis task. IEEE Transactions on Systems, Man, and Cybernetics, SMC-8, 258-271.

Task: Fault diagnosis in graphically displayed network problems

Taxa: Problem solving

Type of Data: Group

Subjects: 8 Engineering graduate students

Training:

1. Trained with and without display aiding

2. Final Test - "Transfer" session of either aided or unaided displays depending on training

Independent Variable(s):

1. Problem size in fault diagnosis using graphically displayed network system as simulated diagnostic task

2. Pacing (forced vs non-forced)

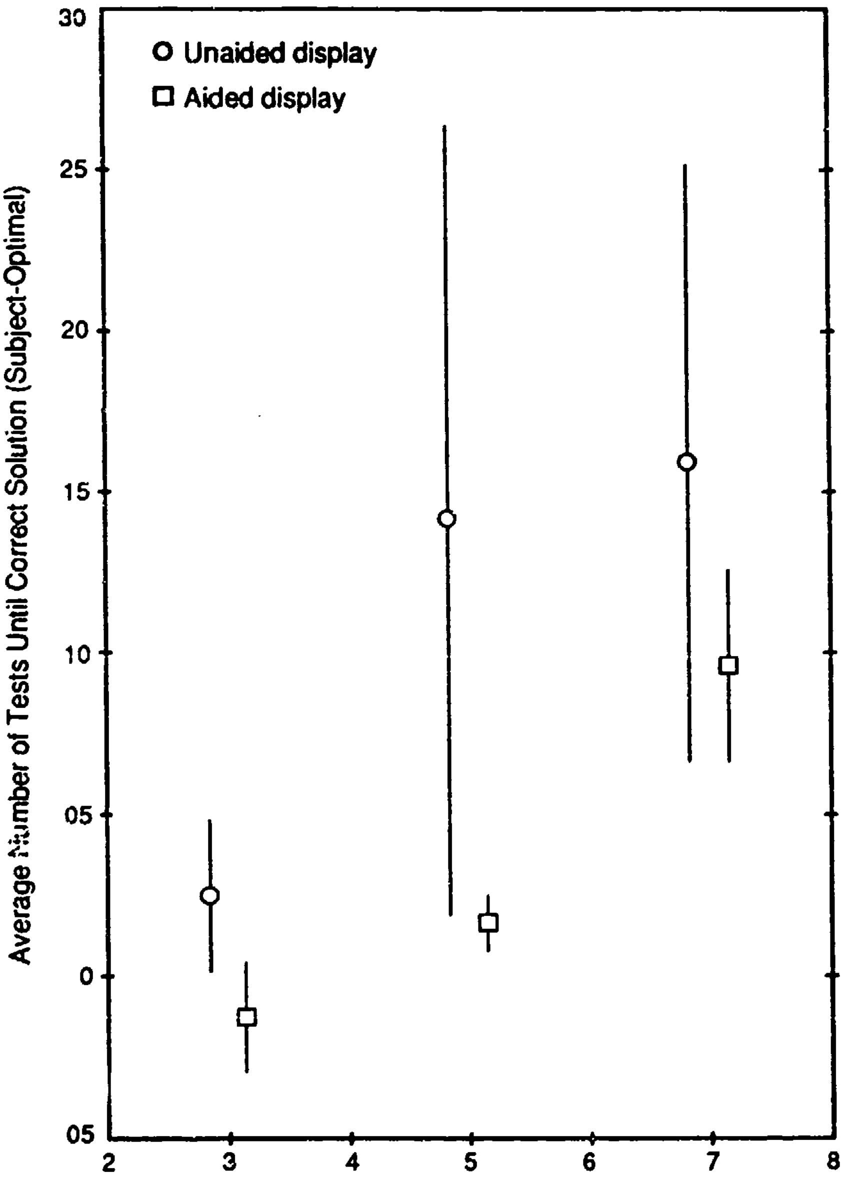
3. Display (computer aided vs non-aided)

Dependent Measure(s):

1. Number of tests until correct diagnosis

2. Percent correct

Abstract: This study examined the design of visual displays used for problem solving tasks. Two experiments were conducted investigating human problem solving performance in the diagnosis of faults in graphically displayed network problems. The effects of problem size, forced-pacing, computer aiding, and training were considered. The results indicated that human performance deviates from optimum as problem size increases. It appears that forced-pacing causes the human to adopt brute force strategies compared with strategies adopted during self-paced situations. Computer aiding greatly lessens the number of mistaken diagnoses by performing the bookkeeping portions of the task.



Average number of tests until correct solution - transfer data-experiment one.

Rouse, 1978 (continued)

Analysis of Average Number of Tests Until Correct Solution (Subject-Optimal) - Training Data - Experiment One

SOURCE	SS	df	Var.	F
Between-Subjects Factors:				
Displays	2.258	1	2.258	2.12
S Nested with Displays	6.128	6	1.021	
Within-Subjects Factors:				
Practice	0.414	1	0.414	2.65
Practice x Displays	0.008	1	0.008	.05
Practice x S Nested with Displays	0.935	6	0.156	
Size	1.825	2	0.913	11.16
Size x Displays	0.433	2	0.217	2.12
Size x S Nested with Displays	0.920	12	0.077	
Practice x Size	1.806	2	0.903	11.16
Practice x Size x Displays	0.015	2	0.008	0.10
Practice x Size x S Nested With Displays	0.984	12	0.082	

Schneider, W., & Fisk (1984). Automatic category search and its transfer. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 1-15.

Task: Visual search

Taxa: Visual/Information processing

Type of Data: Group

Subjects: Experiment 1: 20 Introductory Psychology students
Experiment 2: 6 Introductory Psychology students

Design: Exp. 1: Within-subjects, no repeated measures
Exp. 2: Within-subjects, no repeated measures

Training:

1. Exp. 1: Subjects were provided different number of category exemplars previous to a visual search task

2. Exp. 2: Subjects were trained to detect consistently mapped category words while controlled processing resources were allocated to a variably mapped digit task

Independent Variable(s):

1. Exp. 1: Type of mapping (consistently mapped or variably mapped) used for training of category search task

2. Exp. 2: Number of concurrent tasks (single or dual task)

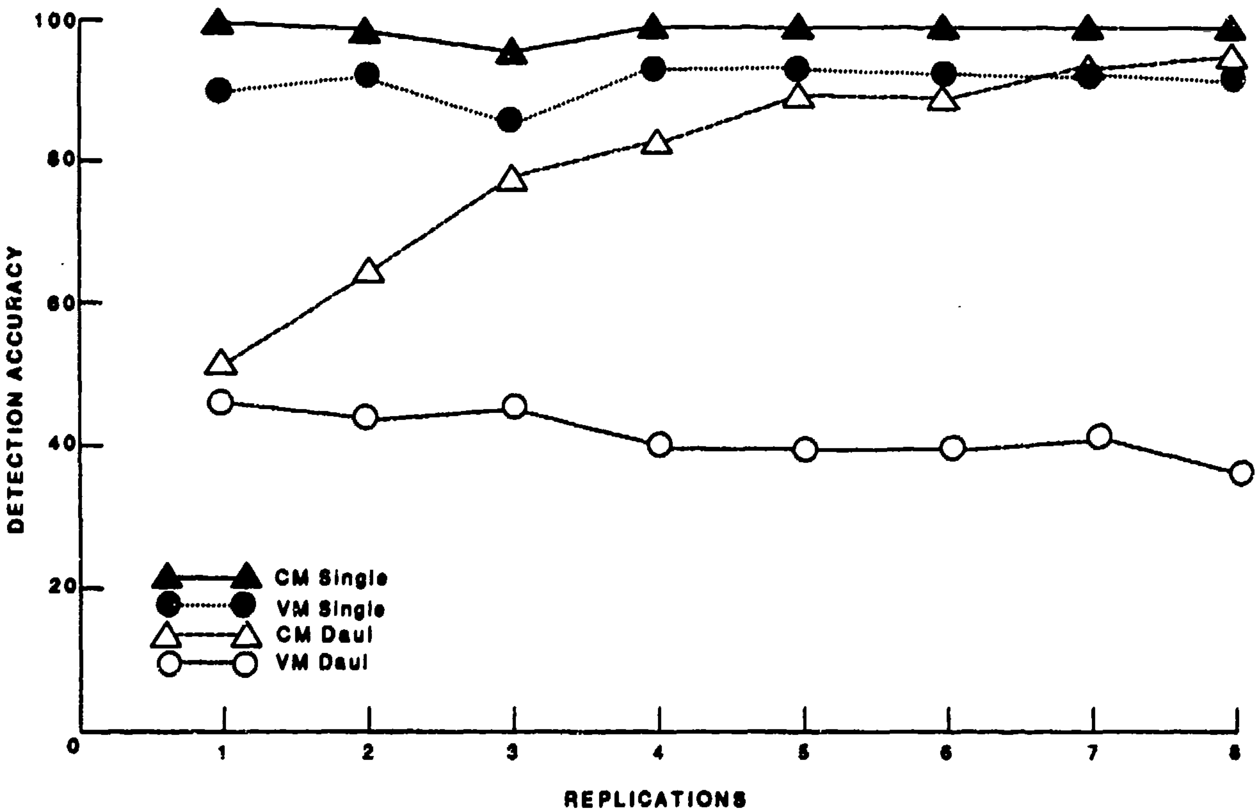
Dependent Measure(s):

1. Exp. 1: Visual search reaction time improvement

2. Exp. 2: Percent improvement of word detection

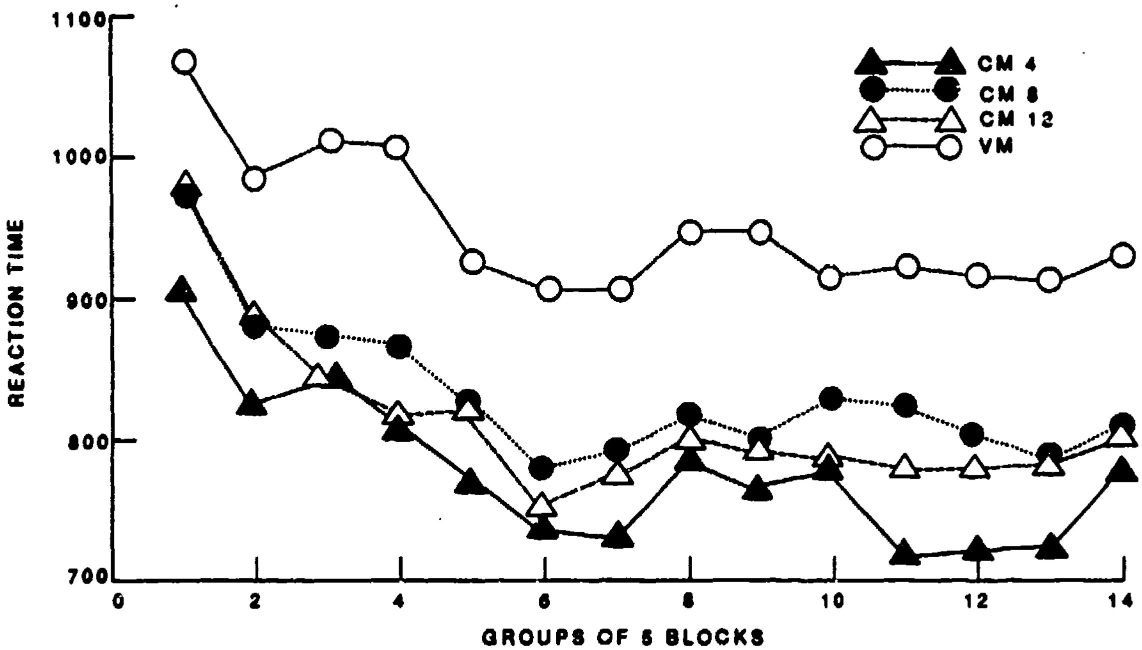
Abstract: Experiments examined practice and transfer effects in consistently mapped (CM) and variably mapped (VM) semantic word search tasks. Experiment 1 examined improvements in reaction time in detecting words from a category as a function of the number of exemplars in the category. All CM conditions showed improvement. Experiment 2 showed that practice reduced resource sensitivity in CM category but did not benefit VM category search.

Schneider et al., 1984



Experiment 2a: Single- and daul-task category detection. (for the first four replications, the category-search conditions varied between trails: the last four search conditions varied between blocks. CM=consistently mapped semantic search. VM=variably mapped semantic search)

Schneider et al., 1984



Experiment 1a: Reaction time as a function of practice. (CM4, CM8, and CM12 refer to training category set sizes of 4, 8, and 12 respectively: reaction time is in milliseconds. Each point represents 35 trials per subject. CM = consistent mapping: VM = varied mapping.)

Scott, P.G., McDaniel, W.C., & Braby, R. (1982). Improved procedures training through use of aids developed from learning guidelines (Navy Technical Report No. 113). Orlando, Fl: Department of the Navy.

Task: Operate a cockpit procedures trainer (CPT)

Taxa: Information processing/Motor

Type of Data: Group

Subjects: 35 newly designated Naval Aviators

Design: Between subjects, no repeated measures

Training:

1. Trained with traditional procedures or with supplemental training materials.

Independent Variable(s):

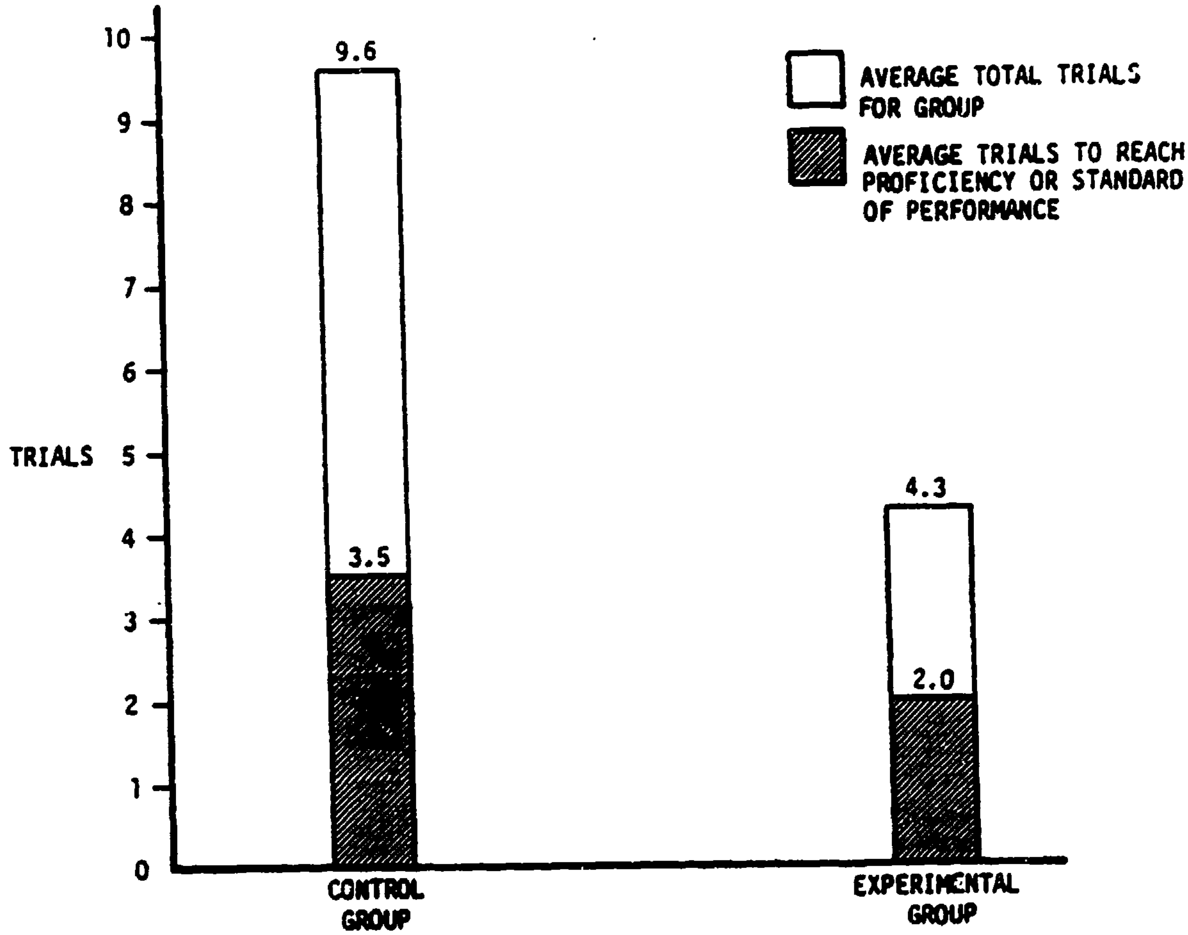
1. SH-3 Aircraft Training aids: traditional or supplemental

Dependent Measure(s):

1. Errors on the cockpit procedures trainer (CPT)

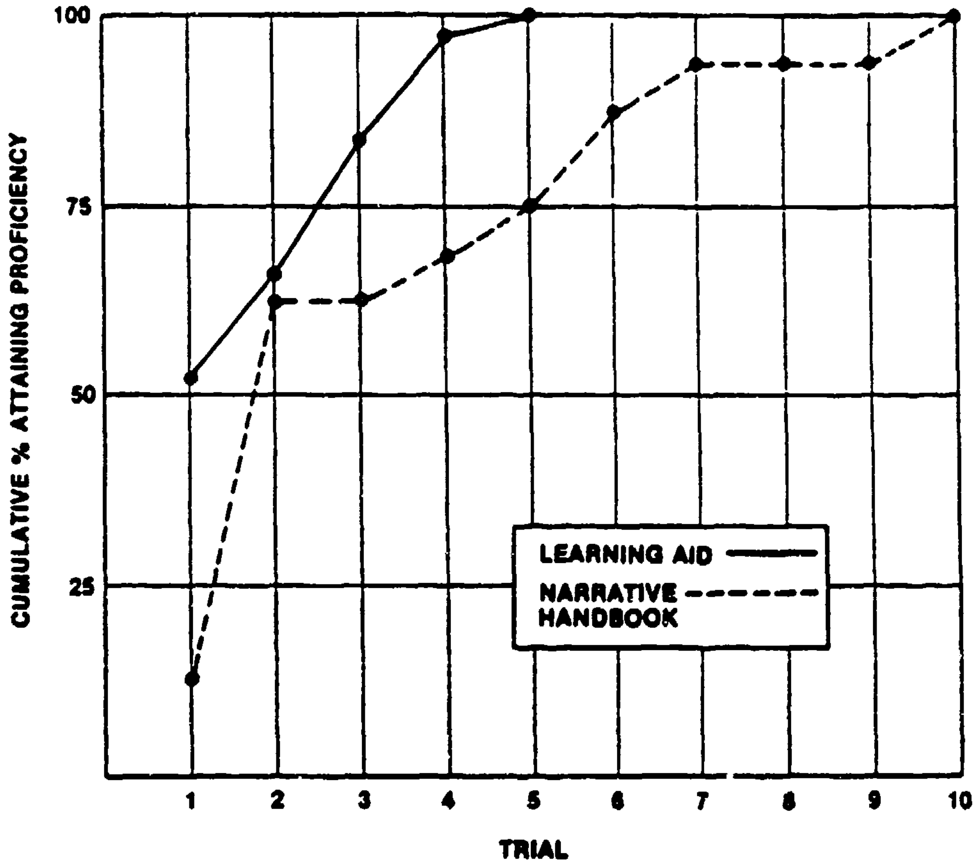
Abstract: This report describes the development of a procedural aid for the SH-3 Aircraft Normal Start Checklist. Subjects either received the training aid or just the traditional instruction for the SH-3 Aircraft. The first measure taken examined students performance on their initial attempt to accomplish the engine start in the CPT (Device 2C44). The second measure examined each students sequence of graded trials in order to determine the point at which the normal start task was performed reliably to NATOPS standards. Results indicated that subjects receiving the procedural aid performed the initial start significantly better than subjects who received traditional instruction. The second measure indicated that subjects receiving the procedural aid demonstrated proficiency significantly sooner than those who received only traditional instruction.

Scott et al., 1982 (continued)



Average Total Trials and Trials to Proficiency for Control and Experimental Groups

Scott et al., 1982 (continued)



Results of Engine Start Field Test Showing Cumulative Subject Proficiency Attainment for Two Types of Training Materials

Sheppard D.J. (1984). Visual and part-task manipulations for teaching simulated carrier landings (NAVTRAEQUIPCEN 81-C-0105-9). Westlake Village, CA: Canyon Research Group, Inc (DTIC No. AD-A153622).

Task: Simulated aircraft carrier landing

Taxa: Motor/Information processing

Type of Data: Group

Subjects: 36 males college students with no flight experience

Design: 2(whole task or part task training) x 2(small or large Fresnel Lens Optical Landing System (FLOLS)) x 3(conventional, rate or command display) factorial design

Training:

1. Subjects were trained under either the part-task condition or the whole-task condition.
2. Subjects were trained using either a small or large FLOLS.
3. Subjects were trained using either a conventional, rate, or command display.

Independent Variable(s):

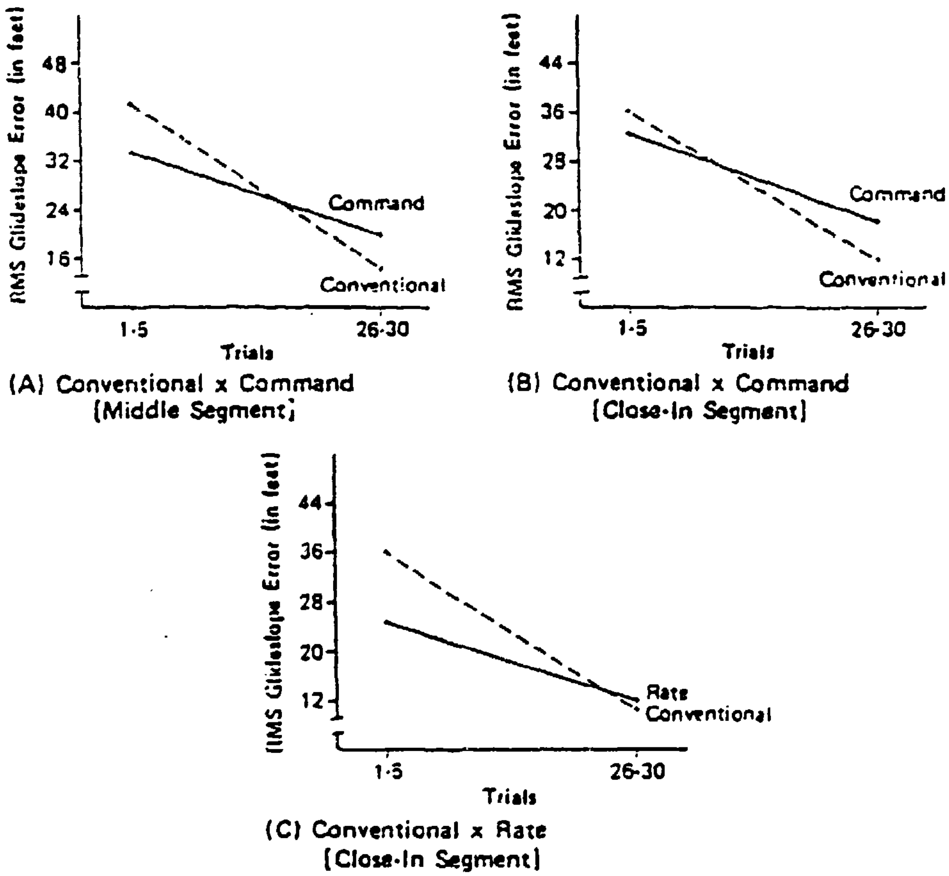
1. Task configuration: part task vs whole task
2. FLOLS type: conventional or vertical bars added
3. FLOLS size: normal size, 1.5 times normal size, 4.5 times normal size

Dependent Measure(s):

1. RMS Glideslope Error

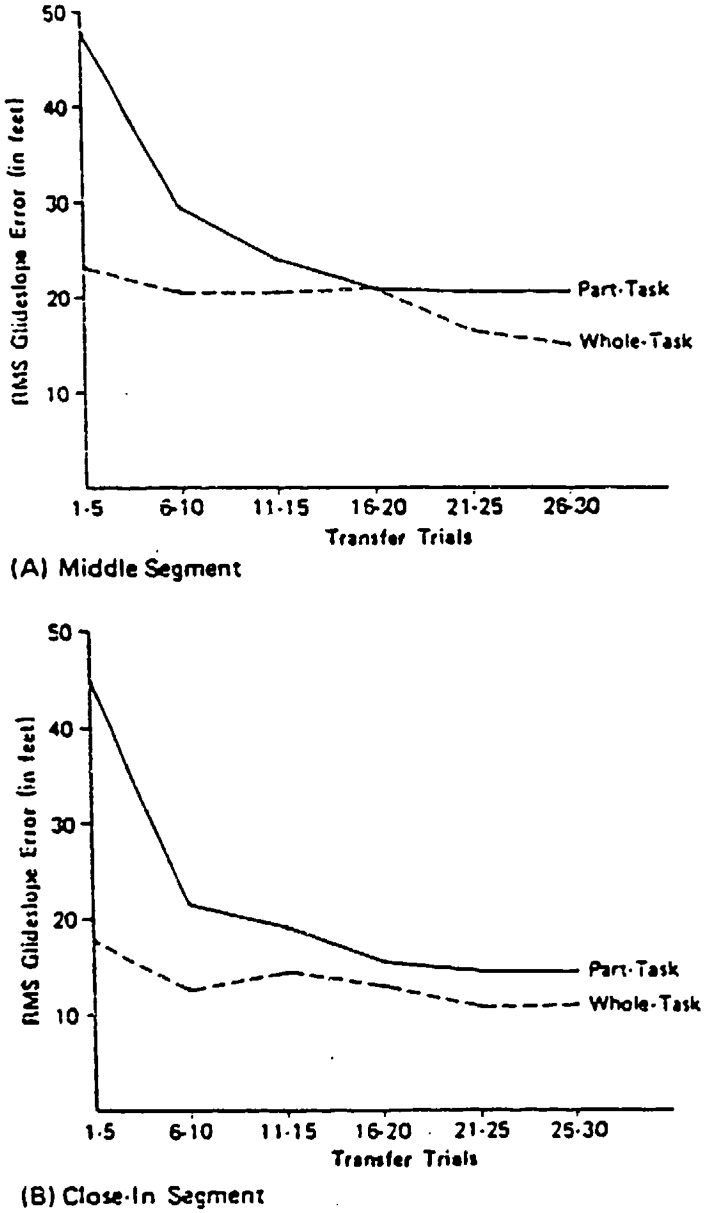
Abstract: This study investigated a segmentation method of part-task training and two types of visual augmentation for teaching simulated carrier landings. The two visual enhancements included a) adding two types of descent rate information and b) enlarging the FLOLS display. The sequence consisted of 30 training trials with instructional feedback under a particular condition followed by 30 test trials with no instructional feedback under the criterion condition. Most importantly, transfer from a large display to a large display had no detrimental effects.

Sheppard, 1984 (continued)



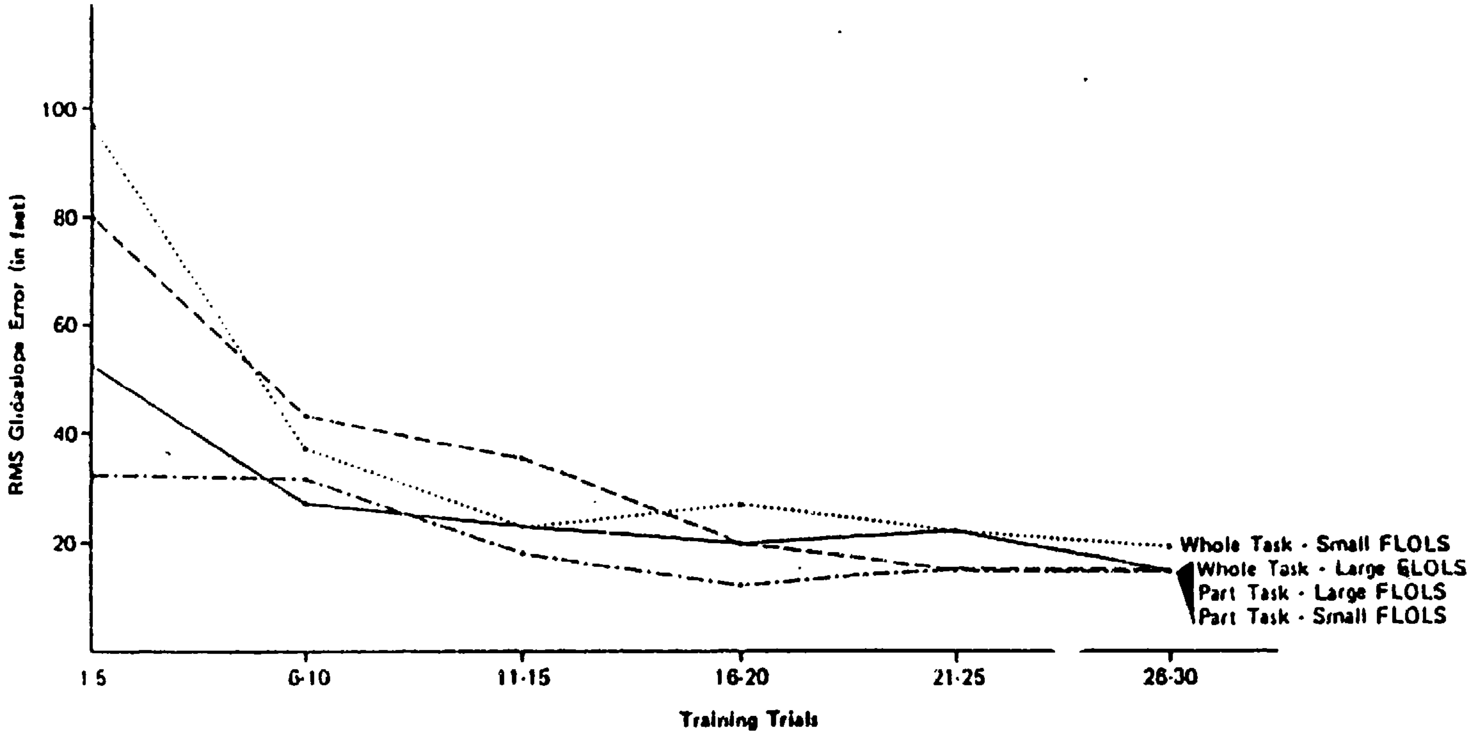
FLOLS type interactions of RMS glideslope error during transfer.

Sheppard, 1984 (continued)



Block x task interactions for RMS glideslope error during transfer.

Sheppard, 1984 (continued)



Block x task x FLOLS size interaction of RMS glideslope error for the middle segment during training.

Simon, C.W., & Roscoe, S.N. (1981). Application of a multi-factor approach to transfer of training research (NAVTRAEQUIPCEN 78-C-0060-6). Westlake Village, CA: Canyon Research Group (AD-A108499).

Task: Tracking of symbols on a plasma-panel display

Taxa: Visual

Type of Data: Group

Subjects: 80 right-handed male non-pilots

Design: 3(control order % acceleration) x 3(display lag) x 3(tracking mode % pursuit) x 3(prediction time) x 3(control gain) x 3(number of training trials) factorial design

Training:

1. Subjects were trained on an aircraft simulator under one of the above conditions

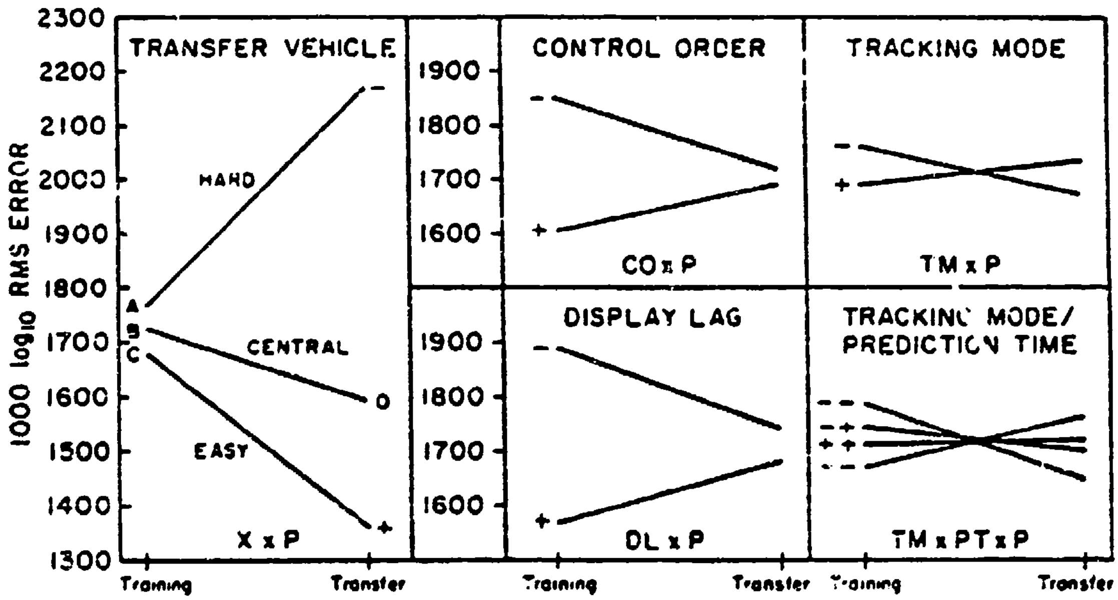
Independent Variable(s):

1. Simulator configurations (variations of the following): Control order, display lag, tracking mode, prediction time, control gain, and number of trials

Dependent Measures):

1. Tracking error as measured on the Visual Technology Research Simulator

Abstract: This study was conducted to establish relationships among training, test, and transfer scores related to the maneuvering of a vehicle, and to determine the relative complexities of response surfaces. Also, this study demonstrated a new transfer research paradigm, and extracted from the available data information that enable the estimation of cost of the effectiveness of simulator characteristics. A horizontal tracking task was used in this study. Six factors were manipulated including vehicle control order, display lag, tracking mode, prediction time, control gain, and number of training trials. Results indicated that the transfer surface appeared less complex than the training surface, the relationship between training and transfer variables was positive but weak, some factors had strong effects in training and weak effects in transfer and vice versa, and transfer was facilitated when the values of certain variables resulted in training conditions that were more difficult than the transfer criterion conditions.



Critical interactions between equipment/training factors and phases.

Weitzman, D.O., Fineberg, M.L., Gade, P.A., & Compton, G.L (1979). Proficiency maintenance and assessment in an instrument flight simulator. Human Factors, 21, 701-710.

Task: To pilot helicopter simulator (Device 2B24)

Taxa: Visual/Motor/Information processing

Type of Data: Group

Subjects: 36 fully qualified combat-ready Army pilots

Design: Repeated measures

Training:

1. Subjects were trained either by aircraft, by simulator, or by both aircraft and simulator.

Independent Variable(s):

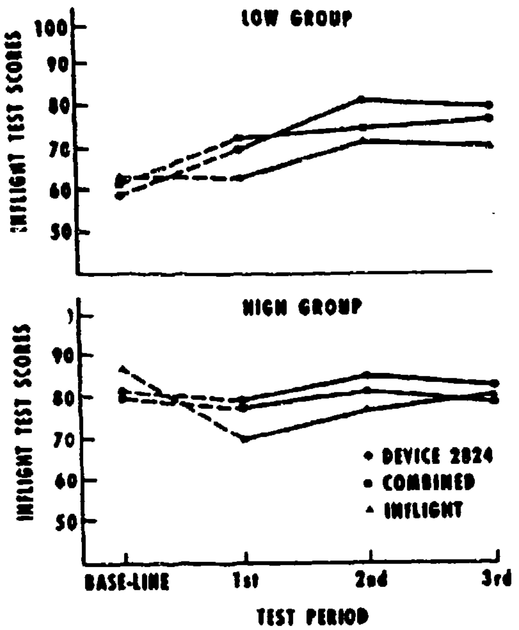
1. Training instrument: Device 2B24 simulator, or the UH-1H helicopter, or both

Dependent Measure(s):

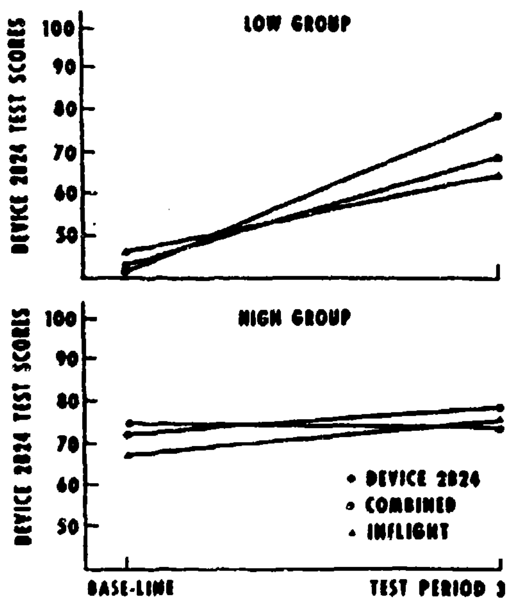
1. Instructor helicopter pilot ratings

Abstract: This study evaluated transfer effects of experienced Army pilots on a high fidelity flight simulator (Device 2B24). Evidence that supported positive transfer was obtained by comparing pilots trained in the simulator with pilots trained in the aircraft (UH-1H) and with pilots trained in both. Also, training in the simulator accurately predicted performance in the aircraft. This suggests that simulators of proven effectiveness can be used both to maintain and assess the proficiency of experienced pilots.

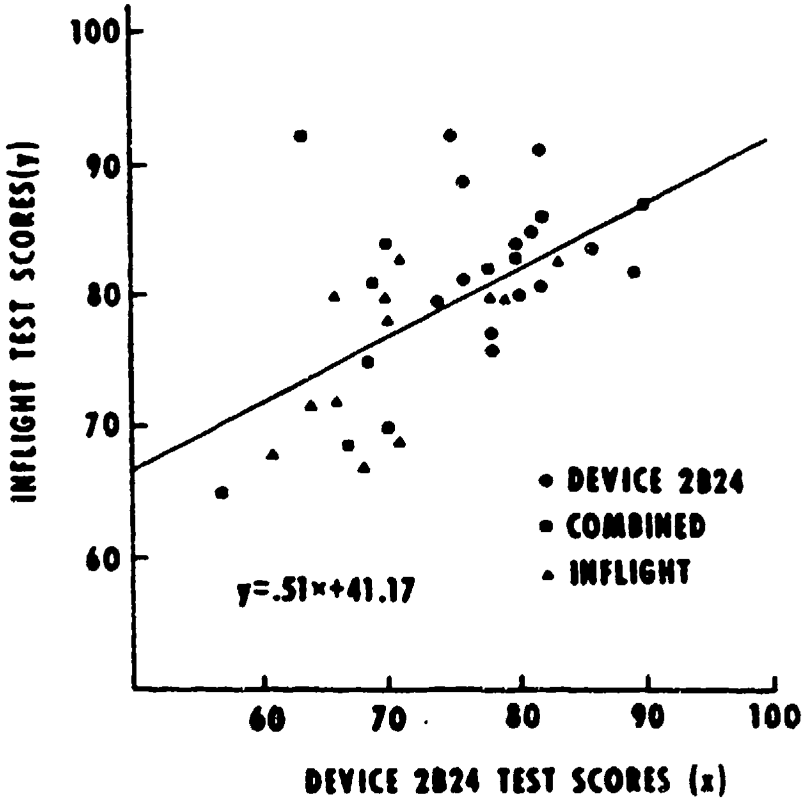
Weitzman et al., 1979 (continued)



Learning curves of checkride performance for the two pilot groups and three training modes.



Simulator performance for the two pilot groups and three training modes.



Inflight test scores plotted against simulator test scores for the last test period. Solid line: least squares, best-fit line.

Wightman, D.C., & Sistrunk, F. (1987). Part-task training strategies in simulated carrier landing final approach training. Human Factors, 29, 245-254.

Task: Simulated carrier landing

Taxa: Motor/Visual

Type of Data: Group

Subjects: 40 male undergraduates

Design: Repeated measures

Training:

1. Subjects were trained on a simulator under conditions of whole task with normal lag, whole task with progressively increased lag, segmented task with normal lag, or segmented task with progressively increased lag.

Independent Variable(s):

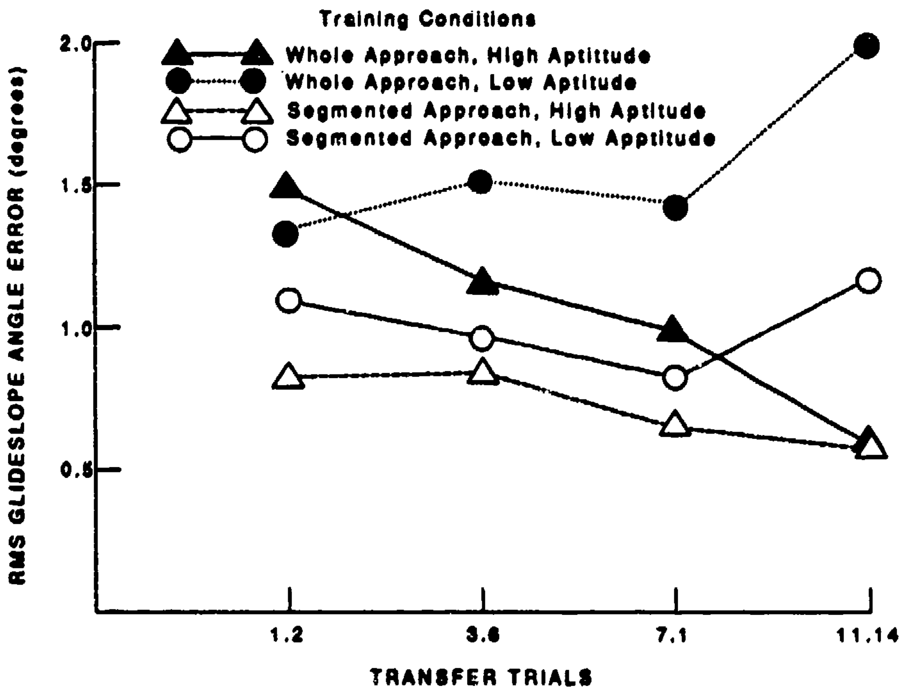
1. Training task: whole task vs segmented task

2. Carrier landing lag: normal lag vs progressive lag

Dependent Measure(s):

1. Root mean squared error (RMS) carrier landing as measured on the Visual Technology Research Simulator

Abstract: This study examined part-task training strategies on transfer to simulator carrier landing. College students were taught carrier landing final approach skills in a simulator under one of three experimental conditions, and were then tested in the simulator on a criterion configuration that was identical to the control training condition. Also, subjects motor-skill aptitude was assessed. Results demonstrated that training under backward procedures of chaining produced better transfer to criterion than did an equal number of training trials on the criterion task itself. An interaction of condition and aptitude revealed that the chaining method was particularly effective for low aptitude subjects.



Graph of Blocks x Task x Motor Skill interaction for transfer trial blocks 1000 feet to the ramp.

Wrisberg, C.A., & Winter, T.P. (1983). Variability of practice and the transfer of training of motor skills (ARI Technical Report 596). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Task: Visual tracking

Taxa: Visual/Motor

Type of Data: Group

Subjects: 288 right-handed students at the University of Tennessee

Design: Repeated measures

Training:

1. Subjects were trained under either a closed skill condition or an open skill condition.

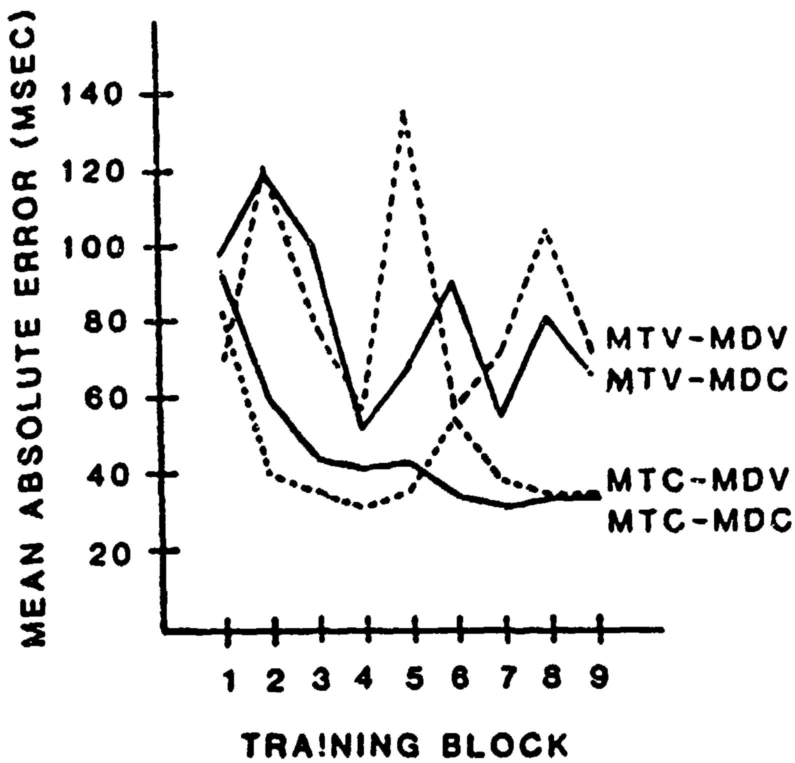
Independent Variable(s):

1. Training condition: open skill (requires varying reaction), closed skill (requires stable reaction)

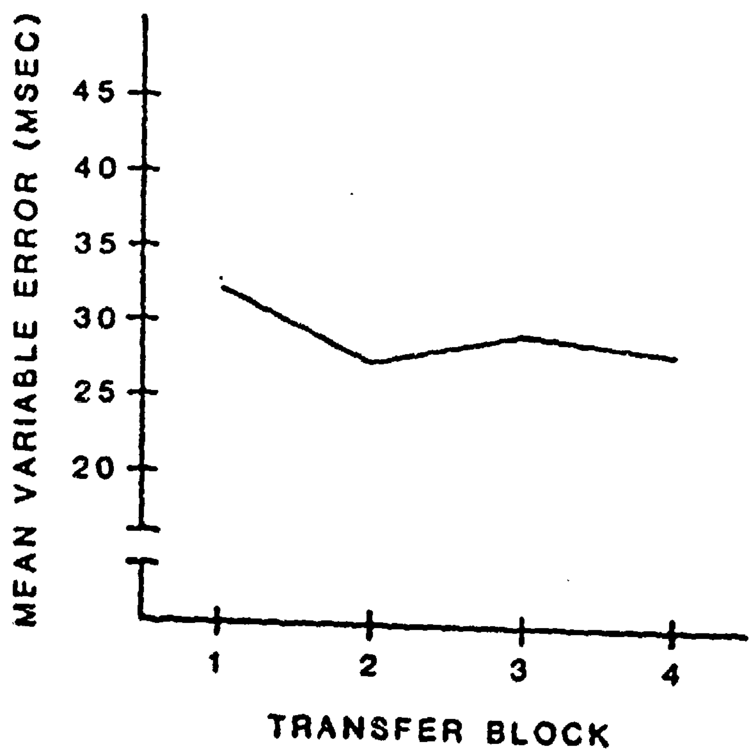
Dependent Measure(s):

1. Absolute error (msec) of arm movement during transfer

Abstract: This study examined whether the amount and variability of practice enhanced transfer to non-practiced tasks of a similar nature. Subjects were trained in either an open skill or open skill situation. Subjects in the open skill situation visually tracked a moving light sequence and attempted to complete a ballistic arm movement coincident with the termination of the light movement. Those in the closed skill condition simply focused their attention on the production of a specific movement on each trial. Variations in movement time and movement distance resulted in higher error for those in the closed skill situation compared with subjects in the open skill condition. However, those in the open skill condition demonstrated transfer performance that was less accurate and less consistent than that of subjects in the closed skill condition.

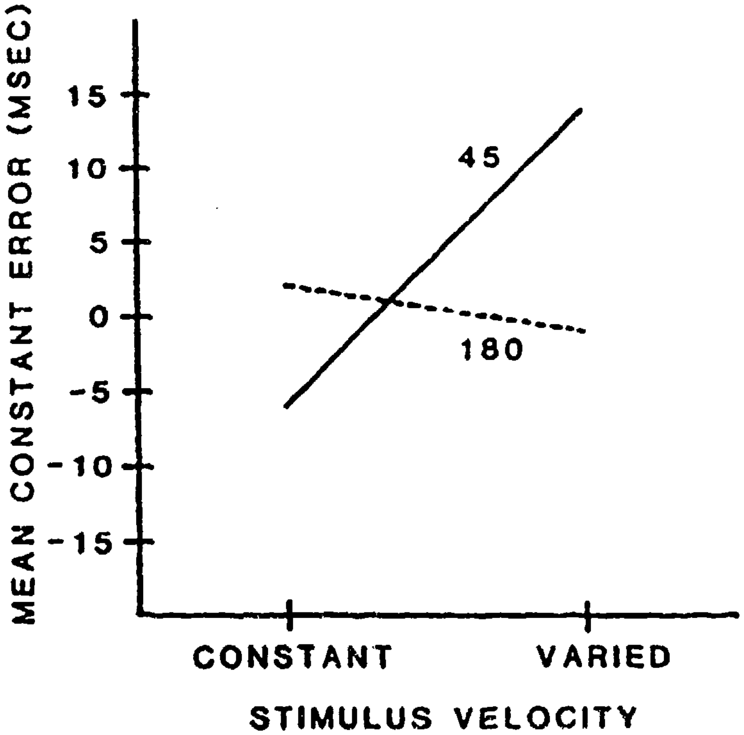


Mean Absolute Error During the Training Phase for the Movement Distance x Movement Time x Blocks Interaction for Closed Skill Subjects Receiving 45 Trials (9 Blocks)



Mean Variable Error for the Four Blocks of Trials During the Closed Skill Transfer Phase

Wrisberg et al., 1983 (continued)



Mean Constant Error During Open Skill Transfer for the Interaction Between Number of Training Trials and Stimulus Velocity Condition During Training

APPENDIX B

REFERENCES FOR RELEVANT RESEARCH ARTICLES

The purpose of Appendix B is to provide a bibliography of articles obtained during the searches that are considered to be relevant to the project. Relevancy of these publications is defined in terms of several criteria including: presence of quantitative data, appropriate statistics and design, use of an adult human population, use of predictor and criterion variables, and use of a task type compatible with one of the nine taxa.

RELEVANT RESEARCH CITATION

Basadur, M., Graen, G.B., & Green, S.G. (1982). Training in creative problem solving: Effects on ideation and problem finding and solving in an industrial research organization. Organizational Behavior and Human Performance, 30, 41-70.

Boreham, N.C. (1985). Transfer of training in the generation of diagnostic hypotheses: The effect of lowering fidelity of simulation. British Journal of Educational Psychology, 55, 213-223.

Briggs, G.E., & Naylor, J.C. (1965). Experiments on team training in a CIC environment (NAVTRADEVCEN 1327-1). Columbus, OH: Human Performance Center (DTIC No. AD-608309).

Card, S.K., English, W.K., and Burr, B.J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. Ergonomics, 21, 601-613.

Cockrell, J.T. (1979). Effective training for target identification under degraded conditions (ARI Technical Paper 358). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A071 085)

Eberts, R.E. (1987). Internal models, tracking strategies, and dual-task performance. Human Factors, 29, 407-419.

Edwards, B.J., Weyer, D.C., & Smith, B.A. (1979). Undergraduate pilot training: Visual discrimination pretraining for landing task (AFHRL-TR-78-78). Brooks Air Force Base, TX: Air Force Systems Command (DTIC No. AD-A068141).

Fisk, A.D., & Lloyd, S.J. (1988). The role of stimulus-to-rule consistency in learning rapid application of spatial rules. Human Factors, 30, 35-49.

Johnson, S.L. (1978). Retention and transfer of training on a procedural task; interaction of training strategy and cognitive style (AFOSR-TR-78-1161). Bolling Air Force Base, DC: Air Force Office of Scientific Research.

Kanfer, R., and Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. Journal of Applied Psychology Monograph, 74, 657-690.

Lintern, G., Roscoe, N.R., Koonce, J.M., & Segal, L.D. (1990).

Transfer of landing skills in beginning flight training. Human Factors, 32, 319-327.

Malloy, T.E., Mitchell, C., & Gordon, O.E. (1987). Training cognitive strategies underlying intelligent problem solving. Perception and Motor Skills, 64, 1039-1046.

Masson, M.E. (1986). Identification of typographically transformed words: Instance-based skill acquisition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12, 479-488.

Miller, E.E. (1975). Instructional strategies using low-cost simulation for electronic maintenance (HumRRO Technical Report HumRRO-FR-WD-TX)-75-20). Alexandria, VA: Human Resource Research Organization.

Morris, N.M., & Rouse, W.B. (1985). The effect of type of knowledge upon human problem solving in a process control task. IEEE, 15, 698707.

Myers, G.L., & Fisk, A.D. (1987). Training consistent task components: Application of automatic and controlled processing theory to industrial task training. Human Factors, 29, 255-268.

Patrick, J., and Haines, B. (1988). Training and transfer of fault-finding skill. Ergonomics, 31(2), 193-210.

Rabbit, P., Cumming, G., and Vyas, S. (1979). Improvement, Learning and Retention of Skill at Visual Search. The Experimental Psychology Society, 31, 441-460.

Roberts, T.L., and Moran, T.P. (1983). The evaluation of text editors: Methodology and empirical results. Communications of the ACM, 26, 265-283.

Rouse, W. B. (1978). Human problem solving performance in a fault diagnosis task. IEEE Transactions on Systems, Man, and Cybernetics, SMC-8, 258-271.

Schneider, W., & Fisk (1984). Automatic category search and its transfer. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 1-15.

Scott, P.G., McDaniel, W.C., & Braby, R. (1982). Improved procedures training through use of aids developed from learning guidelines (Navy Technical Report No. 113). Orlando, FL: Department of the Navy.

Sheppard D.J. (1984). Visual and part-task manipulations for teaching simulated carrier landings (NAVTRAEQUIPCEN 81-C-0105-9). Westlake Village, CA: Canyon Research Group, Inc (DTIC No. AD-A153622).

Simon, C.W., & Roscoe, S.N. (1981). Application of a multifactor approach to transfer of training research (NAVTRAEQUIPCEN 78-C-0060-6). Westlake Village, CA: Canyon Research Group (AD-A108499).

Weitzman, D.O., Fineberg, M.L., Gade, P.A., & Compton, G.L (1979). Proficiency maintenance and assessment in an instrument flight simulator. Human Factors, 21, 701-710.

Wightman, D.C., & Sistrunk, F. (1987). Part-task training strategies in simulated carrier landing final approach training. Human Factors, 29, 245-254.

Wrisberg, C.A., & Winter, T.P. (1983). Variability of practice and the transfer of training of motor skills (ARI Technical Report 596). Alexandria, VA: U.S. Army Research Institute **for the Behavioral and Social Sciences. (AD A142 896)**

APPENDIX C

ANNOTATED BIBLIOGRAPHY OF RESEARCH REVIEWED FOR THIS PROJECT

The purpose of Appendix C is to provide a complete bibliography of the publications that were reviewed during searches for relevant literature. This appendix contains both the publications that were reviewed and at some point rejected and those that were selected as relevant to the study.

ANNOTATED RESEARCH CITATIONS

Adams, J.A., & Hufford, L.E. (1961). Effects of programmed perceptual training on the learning of contact landing skills (NAVTRADEVCEN 297-3). Port Washington, NY: U.S. Naval Training Device Center (DTIC No. AD-264377).

Adams, A.V., & Rayhawk, M.L. (1988). Training effectiveness and cost iterative technique (ARI Research Note 88-57). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences (DTIC No. AD-A197429).

Attneave, F. (1951). Transfer and intercorrelation among three perceptual tasks (HRRC-RN-51-9-P&MS). Lackland Air Force Base, TX: Air Force Personnel and Training Research Center (DTIC No. AD-F760043).

Aushburn, F.B. (1979). Impact of learning styles on air force technical training: Multiple and linear imagery in the presentation of a comparative visual location task to visual and haptic subjects (AFHRL-TR-78-91). Brooks Air Force Base, TX: Air Force Systems Command (DTIC No. AD-A069444).

Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta analysis. Sex Roles, 20, 327-344.

Bahrick, H.P., & Boucher, B. (1968). Retention of visual and verbal codes of the same stimuli. Journal of Experimental Psychology, 78, 417-422.

Bailey, J., Hughes, R.G., & Jones, W.E. (1980). Application of backward chaining to air-to-surface weapons delivery training (AFHRL-TR-79-53). Brooks Air Force Base, TX: Air Force Systems Command (DTIC No. AD-A085610).

Barnett, M.L., Ross, D., Schmidt, R.A., & Todd, B. (1973). Motor skills learning and the specificity of training principle. The Research Quarterly, 44, 440-447.

Basadur, M., Graen, G.B., & Green, S.G. (1982). Training in creative problem solving: Effects on ideation and problem finding and solving in an industrial research organization. Organizational Behavior and Human Performance, 30, 41-70.

Bisseret, A. (1981). Application of signal detection theory to decision making in supervisory control: The effect of the operator's experience. Ergonomics, 24, 81-94.

Boreham, N.C. (1985). Transfer of training in the generation of diagnostic hypotheses: The effect of lowering fidelity of simulation. British Journal of Educational Psychology, 55, 213-223.

Bottger, P.C., & Yetton, P.W. (1987). Improving group performance by training in individual problem solving. Journal of Applied Psychology, 72, 651-657.

Bramley, P. (1973). Equipment recognition training - learning wholes or individual features? Occupational Psychology, 47, 131-139.

Briggs, G.E., & Johnston, W.J. (1967). Team training (NAVTRADEVCEN 1327-4). Columbus, OH: Human Performance Center (DTIC No. AD-660019).

Briggs, G.E., & Naylor, G.E. (1965). Experiment on team training in a CRT-type task environment (NAVTRADEVCEN 1327-1). Columbus, OH: Human Performance Center (DTIC No. AD-608309).

Briggs, G.E., & Johnston, W.J. (1965). Team training research (NAVTRADEVCEN 1327-2). Columbus, OH: Human Performance Center (DTIC No. AD-477965).

Brown, C.J., & Kincaid, J.P. (1981). Assessment of numerical skills of Navy enlisted personnel (Technical Memorandum 81-4). Orlando, FL: Training Analysis and Evaluation Group (DTIC No. AD-A102026).

Buckland, G.H., Monroe, E.G., & Mehrer, K.I. (1980). Flight simulator runway visual textural cues for landing (AFHRL-TR-79-81). Brooks Air Force Base, TX: Air Force Human Resources Lab (DTIC No. AD-A089434).

Buckland, G.H., Monroe, E.G., & Mehrer, K.I. (1979). Flight simulation runway visual textural cues for landing (AFHRL-TR-79-81). Williams Air Force Base, AZ: Air Force Human Resource Laboratory (DTIC No. AD-A089434).

Burns, B., Shepp, B.E. (1988). Dimensional interactions and the structure of psychological space: The representation of hue, saturation, and brightness. Perception & Psychophysics, 43, 494-507.

Card, S.K. (1989). Towards the human interface of the next generation (Xerox Technical Paper P89-00085). Palo Alto, CA: Xerox Palo Alto Research Center.

Card, S.K. (1984). Visual search of computer command menus. In H. Bouma & D.G. Bouwhuis (eds.) Attention and Performance X, Control of Language Processes. Hillsdale, NJ: Lawrence

Erlbaum Associates.

Card, S.K. (1988). Models of working memory (Xerox Technical Paper P88-00117). Palo Alto, CA: Xerox Palo Alto Research Center.

Card, S.K. (1989). Engineering Analyses (Xerox Technical Paper P89-00012). Palo Alto, CA: Xerox Palo Alto Research Center.

Card, S.K., & Moran, T.P. (1980). The keystroke-level model for user performance time with interactive systems. (Xerox Technical Paper SSL-79-1). Palo Alto, CA: Xerox Palo Alto Research Center.

Card, S.K., Pavel, M., & Farrell, J.E. (1985). Window-based computer dialogues. In Human-computer interactions-Interact '84, B. Shackle (ed.). Amsterdam: North-Holland.

Card, S.K. (1979). A method for calculating performance times for users of interactive computing systems (Xerox Technical Paper). Palo Alto, CA: Xerox Palo Alto Research Center.

Card, S.K., English, W.K., & Burr, B.J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. Ergonomics, 21, 601-613.

Card, S.K., Robert, J.M., & Keenan, L.N. (1988). On-line composition of text (Xerox Technical Paper). Palo Alto, CA: Xerox Palo Alto Research Center.

Caro, P.W. (1972). Human performance in transfer of instrument flight training with flight simulator (HUMRRO-PP-7-72). Alexandria, VA: Human Resources Research Organization (DTIC No. AD-743155).

Cockrell, J.T. (1979). Effective training for target identification under degraded conditions (ARI Technical Paper 358). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences (DTIC No. AD-A071085).

Cormier, W.H., Otani, A., & Cormier, L.S (1986). The effects of problem-solving training on two problem-solving tasks. Cognitive Therapy and Research, 10, 95-108.

Crossman, F.W. (1959). A theory of the acquisition of speed-skill. Ergonomics, 2, 153-166.

Curtis, M.J., & Zins, J.E. (1988). Effects of training in consultation and instructor feedback on acquisition o consultation skills. Journal of School Psychology, 26, 185-190.

DeMaio, J., Parkinson, S., Leshowitz, B., & Crosby, J. (1975). Visual scanning: Comparisons between student and instructor pilots (AFHRL-TR-76-10). Brooks Air Force Base, TX: Air Force Systems Command (DTIC No. AD-A023634).

DePaulo, B.M., & Pfeifer, R.L. (1986). On-the-job experience and skill at detecting deception. Journal of Applied Social Psychology, 16, 249-267.

Donderi, D., Seal, S., & Covit, L. (1973). Association learning and visual discrimination. Perception and Psychophysics, 14, 394-400.

Eberts, R.E. (1987). Internal models, tracking strategies, and dual-task performance. Human Factors, 29, 407-419.

Edwards, B.J., Weyer, D.C., & Smith, B.A. (1979). Undergraduate pilot training: Visual discrimination pretraining for landing task (AFHRL-TR-78-78). Brooks Air Force Base, TX: Air Force Systems Command (DTIC No. AD-A068141).

Eialek, H.M., Taylor, J.E., & Hauke, R.N. (1973). Instructional strategies for training men of high and low aptitude (HumRRO-TR-73-10). Alexandria, VA: Human Resources Research Organization (DTIC No. AD-760408).

Ellison, Douglas G., Hill, H. (1947). Wavelength and amplitude characteristics of tracking error curves: Individual differences and learning effects (TSEAA-694-2H). Bloomington, IN: Indiana University (DTIC No. AD-F630370).

Fernandes, K., & Rose, A.M. (1978). An information processing approach to performance assessment: II. An investigation of encoding and retrieval processes in memory (AIR Technical Report No. 2). Washington, DC: American Institute for Research (DTIC No. AD-A062826).

Fico, J.M., & Brodsky, H.S. (1972). The effect of visual and tactual stimulation on learning of abstract forms. Psychonomic Society, 27, 246-248.

Fisk, A.D., & Scheider, W. (1983). Category and word search : Generalizing search principles to complex processing. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 177-195.

Fisk, A.D., & Lloyd, S.J. (1988). The role of stimulus-to-rule consistency in learning rapid application of spatial rules. Human Factors, 30, 35-49.

Fletcher, B.C., & Rabbit, P.M. (1978). The changing pattern of perceptual analytic strategies and response selection with practice in a two-choice reaction time task. Quarterly Journal of Experimental Psychology, 30, 417-427.

Forsyth, G.A., & Brown, D.R. (1968). Stimulus recognizability judgements as a function of the utility of physical dimension in recognition-discrimination problems. Perception & Psychophysics, 3, 85-88.

Forsyth, G.A. (1974). Effects of training in selective attention on perceptual discrimination. Perceptual and Motor Skills, 38, 935-944.

Goldberg, I. (1988). Training effectiveness and cost iterative technique (TECIT) Volume 1: Training effectiveness analysis (ARI Research Note 88-35). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences (DTIC No. AD-A196727).

Grossman, J.D., & Whitehurst, H.O. (1979). The relative effect of multiple factors on target acquisition. Human Factors, 21, 423-432.

Hagman, J.D. (1980). Effects of training task repetition on retention and transfer of maintenance skill (ARI Technical ARI-RR-1271). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences (DTIC No. AD-A101859).

Hagman, J.D., & Rose, A.M. (1983). Retention of military tasks: A review. Human Factors, 25, 199-233.

Haygood, R.C., Parkinson, S.R., Leshowitz, B., & Eddowes, E.E. (1974). Visual and auditory information processing aspects of the acquisition of flying skills (AFHRL-TR-74-79). Brooks Air Force Base, TX: Air Force Systems Command (DTIC No. AD-A007721).

Haywood, K.M. (1983). Response to speed changes in coincidence-anticipation judgements after extended practice. Research Quarterly for Exercise and Sport, 54, 28-32.

Henderson, D.A., & Card, S.K. (1986). Rooms: The use of multiple virtual workspaces to reduce space contention in a window-based graphical user interface. ACM Transactions on Graphics, 5, 211-243.

Hennessy, R.T., Lintern, G., & Collyer, S.C. (1981). Unconventional visual displays for flight training (TR-81-014). Westlake Village, CA: Canyon Research Group, Inc. (DTIC No. AD-A111392).

Hennessy, R.T., Lintern, G., & Collyer, S.C. (1981). Unconventional visual displays for flight training (NAVTRAEQUIPCEN 78-C-0060-5). Orlando, FL: Naval Training Equipment Center (DTIC No. AD-A111392).

Holzman, T.G., Glaser, R., & Pellegrino, J.W. (1976). Process training derived from computer simulation theory. Memory & Cognition, 4, 349-356.

Hord, D.J. (1986). Endogenous ERP components associated with performance in sonar operators: I. Reliability and relation to training (NHRC Report No. 86-35). San Diego, CA: Naval Health Research Center (DTIC No. AD-A182478).

Hosler, W.W., & Richardson, P. (1978). Relation of motor fitness test items to gross motor skills. Perceptual and Motor Skills, 46, 1131-1138.

Johnson, S.L. (1978). Retention and transfer of training on a procedural task; interaction of training strategy and cognitive style (AFOSR-TR-78-1161). Bolling Air Force Base, DC: Air Force Office of Scientific Research (DTIC No. AD-A058966).

Johnson, S.L. (1978). Retention and transfer of training on a procedural task: Interaction of training strategy and cognitive style (Air Force Office of Scientific Research TR-78-1161).

Kalikow, D.N., & Rollins, A.M. (1973). Information processing models and computer aids for human performance (AFOSR Report No. 2654). Arlington, VA: Air Force Office of Scientific Research (DTIC No. AD-777175).

Kanfer, R., and Ackerman, P.L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. Journal of Applied Psychology Monograph, 74, 657-690.

Kennedy, R.S., Jones, M.B., & Baltzley, D.R. (1988). Regions of isoperformance: An interactive software package for trading off training, personnel, and equipment. Orlando, FL: Department of the Air Force.

Kennedy, R.S., Berbaum, K.S., Collyer, S.C., May, J.G., & Dunlap, W.P. (1984). Visual simulation requirements for aircraft aspect recognition at real world distances (NAVTRAEQUIPCEN 81-C-0105-5). Orlando, FL: Naval Training Equipment Center DTIC No. AD-A151040.

King, B.C., Fowler, F.D., & Warner, R.P. (1970). Study, perceptually similar visual environment (NAVTRADEVCEN 69-C--188-1). Orlando, FL: Naval Training Device Center (DTIC No. AD 881976).

Knerr, C.M., Nadler, L.B., & Dowell, S.K. (1983). Comparison of training transfer and effectiveness models. Proceedings of the Human Factors Society, Champaign, IL.

Larichev, O.I., Boichenko, V.S., Moshkovich, H.M., & Sheptalova, L.P. (1980). Modeling multiattribute information processing strategies in a binary decision task. Organizational Behavior and Human Performance, 26, 278-291.

Laxar, K., & Olson, G.M. (1978). Human information processing in navigational displays. Journal of Applied Psychology, 6, 734-740.

Leshowitz, B., Parkinson, S.P., & Waag, W.L. (1974). Visual and auditory information processing in flying skill acquisition (AFHRL-TR-74-103). Brooks Air Force Base, TX: Air Force Systems Command (DTIC No. AD-A009636).

Lindholm, E., & Ruppel, M. (1980). Attention and task complexity as indicated by physiological indices (AFHRL-TR-79-47). Brooks Air Force Base, TX: Air Force Systems Command (DTIC No. AD-A080851).

Lintern, G., Roscoe, S.N., & Sivier, J.E. (1990). Transfer of landing skills in beginning flight training. Human Factors, 32, 319-327.

Lintern, G., & Roscoe, S.N. (1978). Transfer of landing skill after training with visual cues. In: Visual simulation and image realism; Proceedings of the seminar. San Diego, CA.

Lintern, G., Thomley, K.E., Nelson, B.E., & Roscoe, S.N. (1984). Content, variety, and augmentation of simulated visual scenes for teaching air-to-ground attack (NAVTRAEQUIPC-81-C-0105-3). Orlando FL: NTEC (DTIC No. AD-A145219).

Lintern, G., Gavan, A., Sheppard, B., Daniel, J., Parker, C., Donna, L., Yates, D., Karen, E., Nolan, E., & Margaret, D. (1989). Simulator design and instructional features for air-to-ground attack - a transfer study (Essex Report 89A34835). Orlando, FL: Essex Corporation.

Linville, J.M., Liebhaber, M.J., Obermayer, A.H., & Fallesen, J.J. (1989). Computer-mediated group processes in distributed command and control systems: Dyad shared work (ARI Technical Report 842). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences (DTIC No. AD-A211337).

Mackay, W.E., Malone, T.W., Crowston, K., Rao, R., Rosenblitt, D., & Card, S.K. (1988). How do experienced information lens users use rules. (Xerox Technical Paper P88-00138). Palo Alto, CA: Xerox Palo Alto Research Center.

Macpherson, D., Patterson, C., & Mirabella, A. (1989). Application of ARI skill retention model to wheel vehicle maintenance tasks (ARI Research Report 1538). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences (DTIC No. AD-A219684).

Malloy, T.E., Mitchell, C., & Gordon, O.E. (1987). Training cognitive strategies underlying intelligent problem solving. Perception and Motor Skills, 64, 1039-1046.

Martin, E.L., & Cantaneo, D.F. (1980). Computer generated image: Relative training effectiveness of day versus night visual scenes (AFHRL-TR-79-56). Brooks Air Force Base, TX: Air Force Human Resources Lab (DTIC No. AD-A088313).

Martin, E.L. (1981). Training effectiveness of platform motion: Review of motion research involving the advance simulator for pilot training and the simulator for air-to-ground combat (AFHRL-TR-79-51). Brooks Air Force Base, TX: Air Force Human Resources Lab (DTIC No. AD-A095930).

Martin, E.L., & Waag, W.L. (1975). Contributions of platform motion to simulator training effectiveness (AFHRL-TR-78-15). Brooks Air Force Base, TX: Air Force Systems Command (DTIC No. AD-A058416).

Masson, M.E.J. (1986). Identification of typographically transformed words: Instance-based skill acquisition. Journal of Experimental Psychology, 12, 479-488.

McCauley, M.E., Root, R.W., & Muckler, F.A. (1982). Training evaluation of an automated training system for air intercept controllers (NAVTRAEQUIPCEN 81-C-0055-1). Orlando, FL: Naval Training Equipment Center (DTIC No. AD-A123289).

McDonald, B.A., & Whitehill, B.V. (1987). Visual recognition training: Effectiveness of computer graphics (NPRDC Technical Report 87-21). San Diego, CA: Navy Personnel Research and Development Center (DTIC No. AD-A188302).

McGuinness, J., Gouman, J.H., & Puig, J.A. (1982). Effectiveness evaluation for air combat training. Alexandria, VA: Person-System Integration (DTIC No. AD-P000203).

Messick, D.M., & Campos, F.T. (1972). Training and conservation in subjective probability revision. Journal of Experimental Psychology, 94, 335-337.

Mikulecky, L., & Ehlinger, J. (1986). The influence of metacognitive aspects of literacy on job performance of electronics technicians. Journal of Reading Behavior, 18, 41-61.

Miller, E.E. (1975). Instructional strategies using low-cost simulation for electronic maintenance (HumRRO-FR-WD(TX)-75-20. Alexandria, VA: Human Resources Research Organization.

Morris, N.M., & Rouse, W.B. (1985). The effects of type of knowledge upon human problem solving in a process control task. IEEE Transactions on Systems, Man, Cybernetics, 15, 698-707.

Mulhich, D. (1974). Measuring empirical properties of psychomotor skills in different psychological environments. Perceptual and Motor Skills, 39, 1315-1330.

Munro, A., Towne, D.M., & Fehling, M.R. (1981). An experimental system for research on dynamic skills training (Behavioral Technology Laboratories Technical Report No. 96). Redondo Beach, CA: Behavioral Technology Laboratories (DTIC No. AD-A108703).

Munro, A., Cody, J.A., & Towne, D.M. (1982). Instruction mode and instruction intrusiveness in dynamic skill training (Behavioral Technology Laboratories Technical Report No. 3). Redondo Beach, CA: University of Southern California (DTIC No. AD-A121187).

Myers, G.L., & Fisk, A.D. (1987). Training Consistent task components: Application of automatic and controlled processing theory to industrial task training. Human Factors, 29, 255-268.

Nataupsky, M., McCloy, T.M., Bermudez, J.M., & Tirman, V.W. (1980). Effects of criteria on flight simulation: Study 2: Multiple criteria (FJSRL-TR-80-0020). U.S. Air Force Academy, CO: Frank J. Seller Research Lab (DTIC No. AD-A097526).

Naylor, J.C. (1962). Parameters affecting the relative efficiency of part and whole training methods: A review of the literature (NAVTRADEVCEN 950-1). Port Washington, NY: U.S. Naval Training Device Center (DTIC No. AD-275921).

Neves, D.M., & Anderson, J.R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In J.R. Anderson (Ed.) Cognitive skills and their acquisition (pp. 1-55), Hillsdale, NJ: Lawrence Erlbaum Associates.

Newell, A., & Rosenbloom, P.S. (1981). Mechanisms of skill acquisition and the law of practice. In J.R. Anderson (Ed.) Cognitive skills and their acquisition (pp. 1-55), Hillsdale, NJ: Lawrence Erlbaum Associates.

Norman, D.A. (1973). Adaptive training of manual control: Scheme parameters to task parameters (NAVTRAEQUIPCEN 70-C-0218-1). Orlando, FL: Naval Training Equipment Center (DTIC No. AD-764901).

Notterman, J.M., & Tufano, D.R. (1980). Variables influencing outflow-inflow interpretations of tracking performance: Predictability of target motion, transfer function, and practice. Journal of Experimental Psychology: Human Performance and Perception, 85-88.

Ozkaptan, H. (1979). Evaluation of the utility of the theory of signal detection for target acquisition studies (ARI Technical Report 405). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences (DTIC No. AD-A082931).

Patrick, J., & Haines, B. (1988). Training and transfer of fault-finding skill. Ergonomics, 31, 193-210.

Payne, T.A. (1982). Conducting studies of transfer of learning: A practical guide (AFHRL-TR-81-25). Brooks Air Force Base, TX: Air Force Systems Command (DTIC No. AD-A110569).

Prather, D.C. (1971). Trial-and-error versus errorless learning: Training, transfer, and stress. American Journal of Psychology, 84.

Rabbitt, P., Cumming, G., & Vyas, S. (1979). Quarterly Journal of Experimental Psychology, 31, 441-459.

Ratcliff, B., & Siddiqi, J.I.A. (1985). An empirical investigation into problem decomposition strategies used in program design. International Journal of Man-Machine Studies 22, 77-90.

Ratcliff, B., & Siddiqi, J.I. (1985). An empirical investigation into problem decomposition strategies used in program design. International Journal of Man-Machine Studies, 22, 77-90.

Reid, G.B. (1975). Training transfer of a formation flight trainer -- to aircraft formation flying. Human Factors, 17, 470-476.

Roberts, T.L., and Moran, T.P. (1983). The evaluation of text editors: Methodology and empirical results. Communications of the ACM, 26, 265-283.

Rose, A.M., & Fernandes, K. (1977). An information processing approach to performance assessment: I. Experimental investigation of an information processing performance battery (AIR-58500-TR). Arlington, VA: Office of Naval Research (DTIC No. AD-A047299).

Rouse, S.H., & Rouse, W.B. (1982). Cognitive style as a correlate of human problem solving in fault diagnosis tasks. IEEE Transactions on Systems, Man, Cybernetics, 12, 649-652.

Savage, R.C., Williges, B.H., & Williges, R.C. (1982). Empirical prediction models for training-group assignment. Human Factors, 24, 417-426.

Schendel, J.D., Heller, F.H., Finley, D.L., & Hawley, J.K. (1985). Use of weaponeer marksmanship trainer in predicting M16A1 rifle qualification performance. Human Factors, 27, 313-325.

Schendel, J.D., Shields, J.L., & Katz, M.S. (1978). Retention of motor skills: Review (ARI Technical Paper 313). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences (DTIC No. AD-A061338).

Schilling, S., Kaufman, E.J., Schryver, J.C., & D'Amico, A.D. (1985). The application of performance feedback in simulator training: Its effects on the acquisition of shiphandling skills in unfamiliar waterways (CAORF-52-8204-02). Kings Point, NY: National Maritime Research Center (DTIC No. AD-A164472).

Schneider, W., & Fisk, A.D. (1984). Automatic category search and its transfer. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 1-15.

Scott, P.G., McDaniel, W.C., & Braby, R. (1982). Improved procedures training through use of aids developed from learning guidelines (Training Analysis and Evaluation Group Technical Report 113). Orlando, FL: Training Analysis and Evaluation Group.

Sheppard, D.J. (1984). Visual and part-task manipulations for teaching simulated carrier landings (NAVTRAEQUIPCEN 81-C-0105-9). Westlake Village, CA: Canyon Research Group (DTIC No. AD-A153622).

Simon, C.W. (1981). Application of a multifactor approach to transfer of training research (NAVTRAEQUIPCEN 78-C-0060-6). Westlake Village, CA: Canyon Research Group (DTIC No. AD-A108499).

Sneider, W., Vidulich, M., & Yeh, Y. (1982). Training spatial skills for air-traffic control. Proceedings of the Human Factor Society. Champaign, IL.

Sneider, W. (1982). Automatic/control processing concepts and their implications for the training of skills (Report HARL-ONR-8101). Champaign, IL: Human Attention Research Laboratory (DTIC No. AD-A115111).

Spears, W.D. (1982). Process of skill performance: A foundation for design and use of training equipment (NAVTRAEQUIPCEN 78-C-0113-4). Orlando, FL: Naval Training Equipment Center (DTIC No. AD-A136879).

Stabell, C.B. (1978). Integrative complexity of information environment perception and information use: An empirical investigation. Organizational Behavior and Human Performance, 22, 116-142.

Statger, P., & Laabs, G.J. (1977). The effect of divided attention on probe reaction time in multiple-task performance. Canadian Journal of Psychology, 31, 174-215.

Tagatz, G.E., Walsh, M.R., & Layman, J.A. (1969). Learning set and strategy interactions in concept learning. Journal of Educational Psychology, 66, 488-493.

Ulich, E. (1967). Some experiments on the function of mental training in the acquisition of motor skills. Ergonomics, 10, 411-419.

Vidulich, M., Yeh, Y., & Schneider, W. (1983). Time-compressed components for air-intercept control skills. Proceedings of the Human Factors Society. Champaign, IL.

Walsh, J.K. (1973). Effect of visual and tactual stimulation on learning abstract forms: A replication. Bulletin of the Psychonomic Society, 2, 357-359.

Weitzman, D.O., Fineberg, M.L., Gade, P.A., & Compton, G.L. (1979). Proficiency maintenance and assessment in an instrument flight simulator. Human Factors, 21, 701-710.

Westra, D.P. (1982). Simulation and training for aircraft carrier landings - An economical multifactor approach. Human Factors Society Annual Meeting. Seattle, WA.

Westra, D.P. (1982). Investigation of simulator design features for carrier landing: II. In-simulation transfer of training (NAVTRAEQUIPCEN 81-0105-1). Orlando, FL: Naval Training Equipment Center (DTIC No. AD-A124024).

Westra, D.P., Lintern, G., Sheppard, D.J., Thomley, K.E., & Mauk, R. (1986). Simulator design and instructional features for carrier landing: A field transfer study (NAVTRASYSCEN-85-C-0044-2). Washington, DC: Naval Air Systems Command (DTIC No. AD-A169962).

Whitaker, L.A. (1981). Effects of training on response latency and accuracy for a difficult visual discrimination. Perceptual and Motor Skills, 52, 475-478.

Wightman, D.C., & Sistrunk, F. (1987). Part-task training strategies in simulated carrier landing final approach training. Human Factors, 29, 245-254.

Wightman, D.C. (1983). Part-task training in simulated carrier landing final approach training (NAVTRAEQUIPCEN Technical Report IH-347). Orlando, FL: Naval Training Equipment Center (DTIC No. AD-A136670).

Wightman, D.C., & Lintern, G. (1984). Part-task strategies in simulated carrier landing final approach training. Human Factors Society Annual Meeting. San Antonio, TX.

Williams, P.S. (1986). Processing demands, training, and the vigilance decrement. Human Factors, 28, 567-579.

Williges, R.C., & Williges, B.H. (1978). Critical variables in adaptive motor skills training. Human Factors, 20, 201-214.

Wolfe, J., & Jackson, C. (1987). Creating models of the strategic decision making process via participant recall: A free simulation examination. Journal of Management, 13, 123-134.

Wrisberg, C.A., & McLean, E. (1984). Specificity of learning, variability of practice, and the transfer of motor skills training (ARI Technical Report 618). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences (DTIC No. AD-A156794).

Wrisberg, C.A., & Winter, T.P. (1983). Variability of practice and the transfer of training of motor skills (ARI Technical Report 596). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences (DTIC No. AD-A142896).

Yasuara, M., & Kuklinski, T.T. (1978). Category boundary effect for grapheme perception. Perception & Psychophysics, 23, 97-104.

Yates, L.G. (1986). Effect of visualization training on spatial ability test scores. Journal of Mental Imagery, 10, 81-92.

APPENDIX D

ADDDITIONAL DESCRIPTION OF LEARNING CURVE EQUATIONS

The purpose of Appendix D is to provide equations and derivatives used by prominent human learning theorists. This appendix may be used in conjunction with textual discussions of the implementation of learning curves to training and performance issues. Appendix D may also provide a more detailed view of learning curve equations than is provided in the text.

LEARNING CURVE EQUATIONS

Thurstone's Learning Curve Equation

After experimenting with about forty different equations on published learning curves, Thurstone selected a form of the hyperbola taking the form:

$$Y = \frac{L * X}{X + R}$$

in which Y = attainment in terms of the number of successful acts per unit time.

X = formal practice in terms of the total number of practice acts since the start of formal practice.

L = limit of practice in terms of attainment units.

R = rate of learning which indicates the relative rapidity with which the limit of practice is being approached. It is numerically high for a low rate of approach and low for a high rate of approach.

The above equation represents a learning curve which passes through the origin, i.e., starting with a zero score at zero formal practice. However, most learning curves start with some finite score even at the initial performance. For learning curves that do not pass through the origin, the equation is:

$$Y = \frac{L(X+P)}{(X+P)+R}$$

in which P = equivalent previous practice in terms of formal practice units.

Thus, the first equation can be rectified as follows:

$$Y = \frac{L * X}{X + R}$$

$$XY + RY = LX$$

$$X + R = L\left(\frac{X}{Y}\right)$$

The above equation is linear if X/Y is plotted against X. Similarly, the second equation may be rectified if written in the following form:

$$X+(R+P)=L\frac{(X+P)}{Y}$$

which becomes linear when (X + P)/Y is plotted against X. When so rectified, the constants L, R, and P may be determined by several methods. The choice of methods depends on the scatter of the data, the desired accuracy, and the number of curves one must calculate.

<u>Thurstone's Learning Function</u>

The two principle variables in learning are: a) practice and b) attainment. Attainment is an increasing function of practice. Practice may be measured either in terms of repetition or in terms of time devoted to it (effort is assumed to be constant). Repetitions may be counted as successful acts only, or both successes and failures. Attainment can also be expressed in various ways, but for the present purpose it will be expressed in terms of the probability that an act will be counted as successful. This measure of attainment is proportional to the number of successful acts per unit time. An additional assumption is that the rate at which acts are initiated is directly proportional to effort (which is assumed to be constant). Thus far, classifications are as follows:

s = total number of acts that the learner is likely to initiate and that would be credited as successes.

e = total number of acts that the learner is likely to initiate and that would lead to error or failure or loss of time.

p = probability that the act initiate at any moment of time will lead to successful completion.

q = probability that the act initiated at any moment of time will lead to error or loss of time.

Thus, clearly:

$$p=\frac{s}{s+e} \qquad \text{(a)}$$

$$q=\frac{e}{s+e} \tag{b}$$

and

$$ps+pe=s \tag{c}$$

The object of the learner is to reduce the errors e and to raise the proportion p.

Analytically, this can be approached as a probability issue. For the sake of demonstration, let all probable or plausible acts of a learner be represented by white and black balls in a basket. At any moment while the learner is drawing them out, there is a certain probability p that the ball that is drawn out will be white and a certain probability q that it will be black. If it turns out to be black, then there is the probability k that it will be eliminated from the basket and the probability (1-k) that the ball will be returned to the basket.

p = probability that a white ball will be drawn (success)
q = (1-p) = probability that a black ball will be drawn (error)
kq = probability that a black ball will be drawn and eliminated from the basket
q(1-k) = probability that a black ball will be drawn and that this error will be repeated in future attempts

However, when an error is eliminated, obviously the total number of possible errors e is reduced by one. Also, if the probability that a particular draw will give a black ball and that it will be eliminated is kq, then this is also the average number of black balls by which e is reduced per draw. Thus, we are able to write the following differential equation:

$$\frac{de}{dt}=-kq \tag{d}$$

by the earlier equation (b) we have:

$$\frac{de}{dt}=\frac{-ke}{s+e} \tag{e}$$

So far, there has been no mention of the possibility that a successful act has some effect favoring its recurrence. We will assume not only that overly committed errors decrease their probability of recurrence, but also that overtly successful acts increase their probability of recurrence. Thus, following equation (e), we can express a differential equation for successes:

$$\frac{ds}{dt} = +kp \tag{f}$$

which by equation (a) becomes:

$$\frac{ds}{dt} = \frac{ks}{s+e} \tag{g}$$

Here the same constant k is used as was used in equation (e). We can more or less assume that the extent to which a subject profits from his/her errors is about the same as the extent to which he/she profits from his/her successes. Also, we will treat the total number of plausible errors e as a decreasing quantity, and the total number of plausible successes as if it were an increasing quantity. The differential equations (e) and (g) allow us to find a simple relation between the total number of plausible error e and effective number of successes s which the learner has at his/her disposal at any particular moment of practice. Dividing (g) by (e) we have:

$$\frac{ds}{de} = -\frac{s}{e} \tag{h}$$

and consequently:

$$\int \frac{ds}{s} + \int \frac{de}{e} = 0 \tag{i}$$

so that:

$$\log s + \log e = \log m \tag{j}$$

where m is a constant. This may also be written simply as:

$$se = m \tag{k}$$

In effect, the product of the total number of errors and the effective number of successes in the learner's repertoire is a constant. The physical meaning of this constant m is apparently related to the complexity of the learning task. For instance, when the learning process is intrinsically complex for the subject, the product se is large and vice versa.

The constant k is an attribute of the individual learner while the constant m is primarily an attribute of the learning task. If we want to know the relation between practice time t and attainment p, from equation (c) we can write the following differential equation:

$$p\frac{ds}{dt}+s\frac{dp}{dt}+p\frac{de}{dt}+e\frac{dp}{dt}=\frac{ds}{dt} \quad (1)$$

Substituting equations (e) and (g) in (1) we have:

$$\frac{pks}{s+e}+s\frac{dp}{dt}-\frac{pke}{s+e}+e\frac{dp}{dt}=\frac{ks}{s+e} \quad (m)$$

and collecting terms it becomes:

$$(s+e)^2\frac{dp}{dt}=k(pe+s-ps) \quad (n)$$

Thus, the relation between p and t is:

$$\int\frac{(s+e)^2dp}{s-ps+pe}=k\int dt \quad (o)$$

In performing the integration, the variables s and e may be stated first in terms of the individual learning constants k and m and the desired variables p and t. Following from equation (c), this can be done as follows:

$$e=\frac{s-ps}{p} \quad (p)$$

and from equation (k) we have:

$$e=\frac{m}{s} \quad (q)$$

Now, solving equations (p) and (q) simultaneously for s and e, we get:

$$s=\sqrt{\frac{mp}{1-p}} \quad (r\text{-}1)$$

and:

$$e=\sqrt{\frac{m(1-p)}{p}} \qquad \text{(r-2)}$$

The numerator of equation (o) contains the factor $(s + e)^2$ which is evaluated in terms of equations (r-1) and (r-2) as follows:

$$(s+e)=\frac{\sqrt{mp}}{\sqrt{1-p}}+\frac{\sqrt{m(1-p)}}{\sqrt{p}} \qquad \text{(s)}$$

which simplifies as follows:

$$(s+e)^2=\frac{m}{p(1-p)} \qquad \text{(t)}$$

The denominator of equation (o) can also be evaluated in terms of individual learning constants k and m and the desired variables p and t:

$$s-ps+pe=\frac{\sqrt{mp}}{\sqrt{1-p}}-\frac{p\sqrt{mp}}{\sqrt{1-p}}+\frac{p\sqrt{m(1-p)}}{\sqrt{p}} \qquad \text{(u)}$$

which simplifies as:

$$s-ps+pe=2\sqrt{mp(1-p)} \qquad \text{(v)}$$

Equation (o) can be rewritten in a more convenient form by substituting equations (t) and (v) and simplifying:

$$\int\frac{dp}{p^{\frac{3}{2}}(1-p)^{\frac{3}{2}}}=\frac{2k}{\sqrt{m}}\int dt \qquad \text{(w)}$$

Thus, we have a relation between the two variables p and t of the learning curve and the two constants k and m.

In order to facilitate the integration process, one may make the substitution $p = \sin^2\phi$. The limits of p and $\sin^2\phi$ are the same. Therefore, equation (w) becomes:

$$\int\frac{d\phi}{\sin^2\phi\cos^2\phi}=\frac{k}{\sqrt{m}}\int dt \qquad \text{(x)}$$

$$\int csc^2\phi \sec^2\phi d\phi$$

$$=\int(1+\cot^2\phi)(1+\tan^2\phi)d\phi$$

$$=\int(2+\tan^2\phi+\cot^2\phi)d\phi$$

$$=\tan\phi-\cot\phi+C=\frac{kt}{\sqrt{m}} \qquad \text{(y)}$$

But from the substitution it follows that:

$$\tan\phi=\sqrt{\frac{p}{1-p}}$$

and:

$$\cot\phi=\sqrt{\frac{1-p}{p}}$$

Hence equation (w) becomes:

$$\frac{2p-1}{\sqrt{p-p^2}}=\frac{kt}{\sqrt{m}}+z \qquad \text{(z)}$$

in which z is a constant of integration. Finally, this is the desired learning curve equation.

Generalized Power Function (see Lane, 1986)

In its most complete version, the generalized power function appears as follows:

$$T=A+B(N+E)^{-R} \qquad \text{(a)}$$

For the sake of simplicity, all of the following equations will implement the following notational scheme:

N = number of trials, time units elapsed or other index of cumulative exposure or practice

T = time to respond, to perform a task, to complete one unit of output on a given trial, or other measure on which improved performance is indicated by decreasing values.

Y = performance score, output, ratings, or other measure on which improved performance is indicated by increasing values.

A = asymptote. Best possible level of performance. The value of T or Y as N approaches infinity.

B = performance on first trial or output on first trial.

E = prior learning (in trial equivalents). Reflects transfer from prior experience or learning in terms of trials required to attain a presumed entry-level performance.

R = rate variable describing amount of change in Y or T with one unit change in N.

The power function is also encountered in a simpler form with the presumption that A and E are zero:

$$T=BN^{-R} \tag{b}$$

Curves that follow the power law show that the change (improvement) in performance between two trial decreases systematically as the number of trials increases. The rate variable R is a measure of how rapidly improvement drops as a function of practice and also indicates the degree of curvature of the learning curve, the rapidity with which asymptote is reached (if it is reached). The decay under the power law is such that if T (or Y) changes by a given factor over n trials, it will require another n(n - 1) trials for T to change by that factor again (Newell & Rosenbloom, 1981). For these power law curves (as well as most other curves that describe learning), the amount of learning on each trial is a constant proportion of what remains to be learned, thus producing a negatively accelerated curve.

Generalized Exponential (see Lane, 1986)

Please refer to the notational scheme above as needed. The exponential function is encountered as follows:

(a) $$T=A+Be^{-RN}$$

where e is the natural logarithm. As later variants will show, the exponential function can be recast into a variety of forms. The exponential is substantially different from the power law (Newell & Rosenbloom, 1981). The exponential decreases or increases more rapidly since the amount learned on each trial does not decrease as a function of N. Generally, if T decreases by a specific factor in n trials, it takes n additional trials to decrease by that factor again (as opposed to n(n - 1) trials for the power law).

The exponential growth curve or negative exponential usually takes a negatively accelerated form similar in shape to the power function. As shown below, the function increases over time:

(b) $$Y=A[1-e^{-R(N+E)}]$$

where R has the conventional meaning. In this version, the fit is based soley on the asymptote with use of initial value. As shown below, this equation is also sometimes used without the prior experience parameter:

(c) $$Y=A[1-e^{-RN}]$$

A varying notation of the exponential curve employs a different interpretation of the rate variable. As shown in equations (d) and (e) below, the parameter t is a rate constant related reciprocally to R such that large values of t are associated with slower increases in Y:

(d) $$Y=A[1-e^{-(N+E)/t}]$$

(e) $$Y=A[1-e^{-N/t}]$$

A version of the exponential commonly applied in the industrial engineering field is the "time constant" model shown below:

(f) $$Y=B+(A-B)(1-e^{-N/t})$$

The version shown above involves both A and B parameters (see initial notation) and t is the time constant. Note that the quantity (A - B) is the difference between initial performance and asymptote, and thus represents the maximum increase in performance due to learning. The time constant model is primarily applied to the performance of individual operators rather than the total production system.

Logistic (see Spears, 1983; 1985)

Please refer to the notational scheme above as necessary. The logistic curve appears as follows:

$$Y=A/[1+(B-A)E^{-kN}] \qquad \text{(a)}$$

where k is implicitly a function of R such that k = R /[Y (A - Y)], and e is the natural logarithm. Note that in this context, although k is fit as a constant, it is not the same constant rate measure defined for the other curves. It is a measure that varies across trials reflecting a rate that changes in accordance with the amount already learned.

The logistic curve implements four parameters: 1) Beginning Level, 2) Asymptote, 3) , Rate of Learning and 4) Inflection Point. Beginning level is defined as the level of performance before any practice occurs; Asymptote is defined as the limit for performance as practice approaches infinity; and Rate of Learning is defined as the rate variable describing the amount of change in performance with one unit change in practice (see notational scheme above). Finally, and perhaps most importantly, is the Inflection Point component. Spears (1983; 1985) places much emphasis on inflection points as important parameters in evaluating training progress primarily because it is at these points where learning stops increasing and begins to slow down.

APPENDIX E

REFERENCES FOR SUPPLEMENTAL LITERATURE

The purpose of Appendix E is to provide the reader with exposure to additional supporting literature in areas relevant to this project. These publications have been categorized according to general subject area for search convenience.

SUPPLEMENTAL CITATIONS

Experimental and Quasi-experimental Design

Campbell, D.T., & Stanley, J.C. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand McNally & Company.

Sabol, M.A. (1986). Fitting learning curves with orthogonal polynomials (ARI Technical Report 726). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. **(AD A181 148)**

Simon, C.W. (1977). New research paradigm for applied experimental psychology: A system approach (Report No. CWS-03-77). Westlake Village, CA: Canyon Research Group, Inc.

Performance Management

Lane, N.E. (1986a). Issues in performance measurement for military aviation with applications to air combat maneuvering (NTSC TR-86-008). Orlando, FL: Naval Air Systems Command.

Lane, N.E. (1986b). Skill acquisition curves and military training (IDA order no. 5540). Orlando, FL: Essex Corporation.

Mixon, T.R., & Moroney, W.F. (1982). An annotated bibliography of objective pilot performance measures (Report No. NAVTRAEQUIPCEN IH-330). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Simulator Fidelity

Ayres, A., Hays, R.T., Singer, M.J., & Heinickle, M. (1984). An annotated bibliography of abstracts on the use of simulators in technical training (Research Product 84-21). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. **(AD A156 792)**

Puig, J.A., Harris, W.T., & Ricard, G.L. (1978). Motion in flight simulation: An annotated bibliography (Technical Report No. NAVTRAEQUIPCEN IH-298). Orlando, FL: Naval Training Equipment Center.

Spears, W.D., Sheppard, H.J., Roush, M.D., & Richetti, C.L. (1981). Simulator training requirements and effectiveness study (STRES): Abstract Bibliography (AFHRL-TR-80-38). Wright-Patterson Air Force Base, OH: Logistics and Technical

Training Division.

Williges, B.H., Roscoe, S.N., & Williges, R.C. (1973). Synthetic flight training revisited. *Human Factors*, *15*, 543-560.

Task and Behavior Taxonomies

Bownas, D.A., & Cooper, A. (1978). *Review of reported behavior and performance taxonomies* (Institute Report #17). Minneapolis, MN: Personnel Decisions Research Institute.

Hritz, R. (1979). *Behavior taxonomies and training equipment design: A literature review and general model* (AFHRL Contract No. F33615-78-C-0019). Valencia, PA: Applied Science Associates, Inc.

Training Effectiveness Evaluation

Jeantheau, G.G., & Anderson, B.G. (1966). *Training system use and effectiveness evaluation* (Technical Report No. NAVTRADEVCEN 1743.1). Port Washington, NY: U.S. Naval Training Device Center.

Jeantheau, G.G. (1971). *Handbook for training systems evaluation* (NAVTRADEVCEN 66-C-0113-2). Orlando, FL: Naval Training Device Center.

Spears, W.D., Sheppard, H.J., Roush, M.D., & Richetti, C.L. (1981). *Simulator training requirements and effectiveness study (STRES): Abstract Bibliography* (AFHRL-TR-80-38). Wright-Patterson Air Force Base, OH: Logistics and Technical Training Division.

Transfer of Training

Jeantheau, G.G., & Anderson, B.G. (1966). *Training system use and effectiveness evaluation* (Technical Report No. NAVTRADEVCEN 1743.1). Port Washington, NY: U.S. Naval Training Device Center.

Jeantheau, G.G. (1971). *Handbook for training systems evaluation* (NAVTRADEVCEN 66-C-0113-2). Orlando, FL: Naval Training Device Center.

Payne, T.A. (1982). *Conducting studies of transfer of learning: A practical guide*. (Report No. AFHRL-TR-81-25). Brooks Air Force Base, TX: HQ Air Force Human Resources Laboratory.

Valverde, H. H. (1973). A review of flight simulator transfer of training studies. *Human Factors*, *15*, 510-523.