NAVAL TRAINING SYSTEMS CENTER
ORLANDO, FLORIDA

AD-A228 733

CENTER OF EXCELLE
FOR SIMULATION A
TRAINING TECHNO

NAVAL TRAINING SYSTEMS CENTER
ORLANDO, FLORIDA

TECHNICAL REPORT 89-006

A META-ANALYSIS
OF THE FLIGHT SIMULATOR
TRAINING RESEARCH

AUGUST 1990

John W. Jacobs
Carolyn Prince
Robert T. Hays
Eduardo Salas

NAVAL TRAINING SYSTEMS CENTER
HUMAN FACTORS DIVISION
Orlando, FL 32826-3224

DTIC
ELECTE
NOV 02 1990
S E D

W.A. RIZZO, Head
Human Factors Division

H.C. OKRASKI, Director
Research and Engineering
Department

90 11 2 006

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	Approved for public release; distribution is unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S) NAVTRASYSCEN TR-89-006	5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION Naval Training Systems Center	6b. OFFICE SYMBOL (If applicable) Code 262	7a. NAME OF MONITORING ORGANIZATION

6c. ADDRESS (City, State, and ZIP Code) Human Factors Division 12350 Research Parkway Orlando, FL 32826-3224	7b. ADDRESS (City, State, and ZIP Code)

8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Technology	8b. OFFICE SYMBOL (If applicable) Code 222	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

8c. ADDRESS (City, State, and ZIP Code) Office of Naval Research Arlington, VA 22217-5000	10. SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO. 0601152N	PROJECT NO. R000N0	TASK NO. 3358	WORK UNIT ACCESSION NO. DN7080009

11. TITLE (Include Security Classification)

(U) A Meta-Analysis of the Flight Simulation Training Research

12. PERSONAL AUTHOR(S)
Jacobs, J. W.*, Prince, C., Hays, R. T., and Salas, E.

13a. TYPE OF REPORT Final	13b. TIME COVERED FROM 1986 TO 1989	14. DATE OF REPORT (Year, Month, Day) 90,08,01	15. PAGE COUNT 55

16. SUPPLEMENTARY NOTATION *Appointed as a post-doctoral fellow under the Research Participation Program administered by Oak Ridge Associated Universities through an agreement between the U.S. Dept. of Energy and NAVTRASYSCEN.

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Meta-analysis, Flight Simulation, Training Effectiveness, Training System Design

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

A meta-analysis of flight simulation research was conducted to identify important characteristics associated with the effectiveness of simulator training. A total of 247 articles, research reports, and technical reports were located, of which 26 had sufficient information for statistical meta-analysis; 19 involved jet pilot training and 7 involved helicopter pilot training. The major finding was that use of simulators consistently produced improvements in training for jets (relative to aircraft training only). No conclusion about simulator effectiveness for helicopter training could be made due to the small number of experiments available for this analysis. Use of motion cuing added little to the training environments for jets, and may have even detracted from training for some tasks. For helicopters, the effects of motion cuing were not able to be analyzed because of an insufficient number of experiments. Conclusions concerning the benefits of motion cuing for both types of aircraft were considered highly tentative due to methods used when conducting motion-related experiments. In general, training outcomes appear to be influenced by the type of task and the amount and type of training. An agenda for future research is provided.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION
22a. NAME OF RESPONSIBLE INDIVIDUAL C. Prince	22b. TELEPHONE (Include Area Code) (407) 380-4831	22c. OFFICE SYMBOL Code 262

DD FORM 1473, 84 MAR   83 APR edition may be used until exhausted.
All other editions are obsolete

*U.S. Government Printing Office: 1986—607-047

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Scott I. Tannenbaum of the State University of New York at Albany, Dr. Terry L. Dickinson of Old Dominion University, and Dr. John E. Mathieu of Pennsylvania State University for their assistance on statistical procedures and analysis. Special thanks are also extended to numerous Naval aviators and NAVTRASYSCEN engineers who provided subject matter expertise on flight tasks and training simulator characteristics.

```
Accession For
NTIS  GRA&I        X
DTIC TAB           ☐
Unannounced        ☐
Justification

By
Distribution/
Availability Codes
        Avail and/or
Dist      Special

A-1
```

DTIC
COPY
INSPECTED
6

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

PROBLEM

The aviation training research examining the effectiveness of simulator training has been so diverse that the results of the individual investigations have been difficult to combine. Traditional narrative reviews have produced inconsistent conclusions. This has resulted in an inability to derive specific guidance for training design from the accumulated research.

OBJECTIVE

The objective of this review was to apply recent advances in data integration to the aviation simulator training effectiveness research, in order to identify those characteristics that have an impact on training outcomes. Areas of interest included: 1) imput variables (task equipment, task requirements, and trainee characteristics); and 2) throughput variables (simulator design and training context).

APPROACH

A total of 247 journal articles and technical reports that addressed aviation training were located. From this base, experiments which included training transfer to the actual equipment were selected. A quantitative review approach (collectively referred to as meta-analysis) was applied to those experiments that reported the information required for the statistical analysis. A total of 26 experiments (19 involving jet aircraft and seven involving helicopters) were included in the final meta-analysis.

CONCLUSIONS

The research reviewed for this analysis demonstrated that the use of simulators consistently produced superior training for jet pilots (relative to aircraft-only training). Since the analysis included such a small number of helicopter experiments, no conclusion on the training effectiveness of the helicopter simulators could be drawn. Motion cuing was found not to add significantly to the training for jet pilots, and in some cases, may have detracted from the training. The conclusions concerning the training outcomes for motion-based simulators were considered highly tentative due to methods that had been used when the motion-related experiments were conducted. There were too few experiments comparing training in motion based simulators to training with no motion for helicopter pilots for analysis to be done. In general, training outcomes appear to be influenced considerably by the type of task trained and the amount and type of training given.

Several specific training variables were examined. The findings from these areas are as follows:

7

Task Equipment

The outcomes of the experiments involving the training of jet pilots were different from those involving the training of helicopter pilots. Results differed in both size and pattern of training outcomes. Jet experiments consistently found simulator training combined with aircraft training to be better than training in the aircraft alone. The findings from similar helicopter experiments were less consistent, and only slightly favored simulator training combined with aircraft training over aircraft training alone.

An insufficient number of helicopter experiments (total N=7) precluded any in-depth analysis involving this type of aircraft. Therefore the results of the meta-analysis are specific to jet aircraft training involving recent Undergraduate Pilot Training (UPT) graduates or current trainees with little or no experience in a simulator or in a jet aircraft.

For jets, the overall training effect for all tasks trained was positive and robust. Over 90 percent of the experimental comparisons favored the simulator and aircraft trained group over the aircraft-only trained group. On the average, subjective performance measures (e.g., instructor ratings) were more sensitive to training effects, and produced greater results than those obtained with objective measures (e.g., instrument readings). As training for both groups progressed and reached the point where it was conducted solely in the aircraft, differences between the groups diminished.

Task Requirements

Certain tasks were more effectively trained in the simulator than others. For jets, when simulators were used for the training of takeoff, approach (to landing), and landing (excluding carrier landings) tasks, the training effects were greater than they were for the combination of all tasks.

Trainee Characteristics

Only two trainee characteristics were identified as likely to have an effect on training results, flight experience and UPT grades. These differences in trainees were rarely studied. When there was concern that these differences might affect training in any single experiment, an effort was made to compose each of the trainee groups with equal amounts of experience or similar grades.

Simulator Design

For jet training, motion cuing was found to add nothing to the simulator training effectiveness, and in some cases, may have taken away from the training value of the simulator. However, this finding may not be truly representative of the effectiveness of motion-based training since: 1) there was a lack of periodic calibration of the motion cuing systems; and 2) the

results were based on all tasks combined. The positive effects of motion for any one task may have been masked by the negative effects of motion for another task.

Training Context

The average effectiveness for training programs where trainees were allowed to progress based on a demonstrated proficiency was greater than for training programs where all trainees proceeded at the same pace. Information on other aspects of the training context, such as the use of instructional features and the provision of feedback was seldom reported and could not, therefore, be analyzed.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS - (Continued)

LIST OF ILLUSTRATIONS

LIST OF TABLES

TABLE OF CONTENTS - (Continued)

INTRODUCTION

Since the introduction of flight simulators to aviation training, questions about the most effective designs of the resulting systems have proliferated. There long has been the expectation that answers to these questions are to be found in the extensive research that has evaluated the use of flight simulators in training. However, individual research results have been mixed, and narrative reviews of the research domain have failed to provide specific training system requirements for producing the most favorable training outcomes.

Several reasons may be cited for the disparity of results in this domain. For example, Caro, in 1973, noted that the primary focus of research efforts had been to determine the effectiveness of a particular simulator within a given training program, rather than to manipulate various training system elements, in order to develop general design principles. Nearly ten years later, Waag (1981), in reviewing more recent research, made the same observation.

The narrow research focus for experimentation within this area was adopted, in part, from practical requirements. The cost of a full mission simulator, and the operational aircraft required for assessing transfer of training (TOT) dictate that most research in aviation training must occur in an on-going training program. Conducting research in this setting necessitates that the research be secondary to safety factors and to the training program. This almost inevitably interferes with experimental control (Osborne, Broyles, & Quick, 1985) and dictates that the research will be designed to answer the one question that is most pressing for the training organization: Does this simulator train?

Another factor that helps shape research objectives rises from the underlying assumption regarding the role of the simulator within the training program (Eddowes & Waag, 1980). Due to efforts aimed at saving on the cost of operating the aircraft, the simulator, in many cases, has been considered a substitute aircraft rather than a training device. This view places emphasis on the similarity of the simulator to the operational equipment, and tends to de-emphasize the investigation of other elements that may have an impact on training. In support of this viewpoint, there is theory and research suggesting that increasing the common elements that exist between the training and operational environments increases transfer of training (Osgood, 1949; Thorndike, 1903).

According to Eddowes and Waag (1980), the simulator can alternatively be viewed as a teaching tool, and its effectiveness can be improved in ways other than by increasing its physical similarity to the operational equipment. This assumption is supported by mounting experimental results indicating that positive training outcomes may be realized using simulators that do not have a high physical resemblance to the operational aircraft (see e.g., Caro, Corley, Spears, & Blaiwes, 1984). Prophet and Boyd (1970) demonstrated that

15

procedural training could be just as effective when trainees practiced on a simple mockup of a cockpit as when they trained in a sophisticated computerized trainer. Finally, there is convincing evidence indicating that the effectiveness of a flight simulator varies according to the training method used (see e.g., Bailey, Hughes, & Jones, 1980).

The term commonly used to describe the degree of realism between the simulated and operational environments is fidelity. Fidelity has been used in a variety of contexts and has been given a number of definitions (Hays, 1980). For this review, the term fidelity is based on the definition presented by Hays and Singer (1989).

> Simulation fidelity is the degree of similarity between the training situation and the operational situation which is simulated. It is a two dimensional measurement of this similarity in terms of: (1) the physical characteristics, for example, visual, spatial, kinesthetic, etc.; and (2) the functional characteristics, for example, the informational, and stimulus and response options of the training situation (Hays & Singer, 1989, p. 50).

As this definition makes clear, the realism of simulation is a complex concept. The majority of research has examined only the physical aspects of this concept, rather than the functional aspects.

With so much variance in experimental objectives and in training orientations, there are no individual experiments with aviation training devices that can answer questions of general interest for training. Integration is necessary to determine if this body of research can produce guidance for training system design and for future research.

RESEARCH INTEGRATION

The traditional form of review, the narrative review, has been unequal to the task of integrating the results from diverse experiments. Reviewers of aviation training research are required to make a large number of judgements in combining the information that provides a summary of the research area. These judgments include comparisons of issues associated with: experimental control, training tasks, level of trainees, length of training programs, and the relative value of one reported statistic over another. The number of decisions to be made is large, and the reviewer has no guidance in making these decisions. Furthermore, since he/she is not required to document in the review how the decisions were made, it is difficult for the reader to assess the relative value of two different reviews. To correct for apparent shortcomings inherent in the narrative review method (see Jackson, 1980), several quantitative review techniques, collectively known as meta-analysis, have been developed (Glass, McGaw & Smith, 1981; Hunter, Schmidt & Jackson, 1982; Rosenthal, 1978).

META-ANALYSIS

Meta-analysis seeks to aggregate and transform individual research outcomes into a common effect size metric (e.g., d or r), then to compute a mean value across experiments to obtain a good estimate of the population value (Glass et al., 1981). While the techniques involved may vary, meta-analytic reviews are becoming an increasingly popular tool for social scientists (Bangert-Drowns, 1986). One important advantage of meta-analysis over the narrative review method is the explicit information provided on the decision processes used by the reviewer (Mullen & Rosenthal, 1985).

Mullen and Rosenthal (1985) note that the combination of some research characteristics relies heavily on the subjective decision-making processes of the reviewer. They presented several decisions that may critically affect the outcome of a meta-analytic review, including: 1) the choice, coding, and use of research characteristics; 2) decisions inherent in the data/retrieval reconstruction process; and 3) methods used to summarize results across experiments. The approach they recommend for appropriately dealing with subjective decision processes is to, "...make explicit the rationale behind, and the procedures underlying, those coding schemes used," (p. 18), and to impart reliability to the decision-making process by using several coders.

At least two different approaches to meta-analysis have emerged. These approaches reflect different philosophies concerning variation in effect sizes (Mathieu & Tannenbaum, 1983; see also Dickinson, Hassett, & Tannenbaum, 1986). The first, advocated by Glass et al. (1981), assumes that the variability in effect sizes within a given domain is due to moderator variables. For example, training effectiveness could be modified by characteristics of the simulator, the trainees, the instructor, or other moderator variables. According to this approach, effect sizes are regressed upon the moderator variables of interest and the resulting outcome is used to explain differences between the research effect sizes.

Another approach to meta-analysis is based on the work by Hunter et al. (1982). This approach differs from the Glassian (1981) approach, in that it is more conservative with regard to moderator variables. Specifically, Hunter et al. (1982) caution that variation of effect sizes may partially result from such artifacts as: 1) sampling error; 2) measurement unreliability; and 3) range restriction. Their approach advocates correcting for these artifacts prior to the search for valid moderator variables. It follows that if sufficient unexplained variability inherent in the effect sizes remains after removing error variance from the above three sources, then a search for moderator variables is warranted. In general, the Hunter et al. (1982) approach is more conservative than the Glassian (1981) approach because it minimizes the likelihood of incorrectly inferring a valid moderator exists. This review incorporates meta-analytic procedures advocated by Hunter et al. (1982), although formulas from Glass et al. (1981) were used to derive effect size values within a given experiment.

In sum, several factors have made it difficult to accurately determine what variables affect flight simulator training outcomes, and to what degree they do so. One important factor is the overly narrow focus of individual experiments that make up the research in this area. Traditional narrative reviews of this domain have failed to extract the information that would allow specification of training principles. An alternative to the narrative review is meta-analysis, which employs quantitative review techniques.

OBJECTIVES

In light of the above, the objectives for this research were to conduct meta-analysis in order to: 1) identify variables that affect flight simulator training outcomes; 2) identify information gaps in the literature (i.e., variables of interest that have yet to be systematically evaluated); and 3) provide direction for future research. Satisfaction of these objectives will contribute to the larger objective of improved flight training for military pilots.

METHOD

In order to analyze research reports in the aviation training area, with regard to variables that may affect training outcomes, a conceptual meta-model was developed. The meta-model is presented in Figure 1, and includes the two broad areas thought to have an impact on the magnitude of training outcomes: 1) input variables, such as the requirements of the task, task equipment, and trainee characteristics; and 2) throughput variables, including simulator design and training context. These areas were selected from training variables that were most consistently identified as important in various reviews of the simulator training research cited below. With the exception of trainee characteristics, the areas of interest in the meta-model are consistent with those specified by Wheaton et al. (1976). Trainee characteristics were added to the present review because there is a growing interest in the area of individual differences and how such differences may affect training (see e.g., Hogan, Arneson, & Salas, 1987; Jones, Kennedy, Turnage, Kuntz, & Jones, 1986).

INPUT	THROUGHPUT	OUTPUT
	...> Simulator Design ...	
Task Equipment		Training Outcomes
Task Requirements		...> Trainee Performance
Trainee Characteristics		
	Training Context	
	...> -Training Type..........	
	-Performance Measurement	
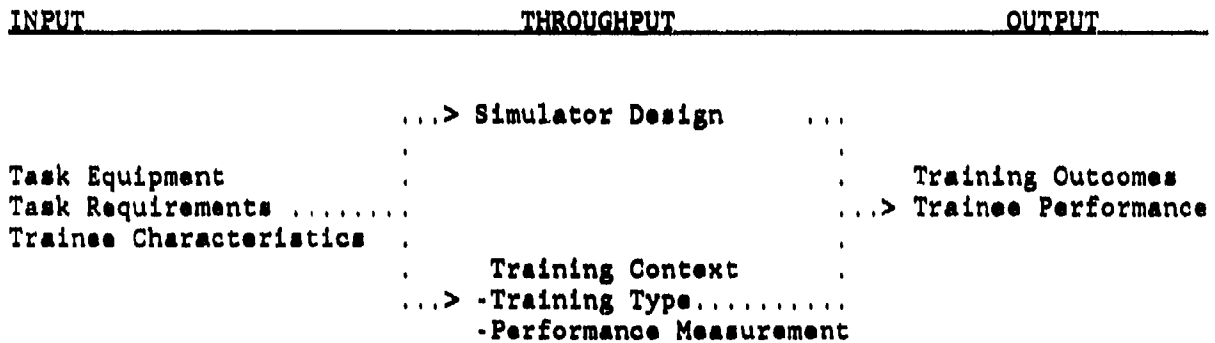
Figure 1. The Meta-Model for Simulator Fidelity

LITERATURE SEARCH

Experiments within the domain of simulator fidelity and training effectiveness were first identified through An Annotated Bibliography of Abstracts on the Use of Simulators for Technical Training (Ayres, Hays, Singer, & Heinicke, 1984). This document was compiled after literature searches for the years 1957 to 1982. Additional searches were conducted for articles for the years 1982 to 1986. Other experiments were located through the reference lists in the obtained articles and in a published search from the U.S. Department of Commerce on Flight Simulator Training (December, 1983 to November, 1985). An open letter was given to all attendees of the Fourth Annual Flight Simulation Update Conference - 1988, requesting recent articles, published or unpublished, that addressed transfer of training in aviation. In addition, individual researchers were contacted and asked for information on relevant research that may not have been included in the published domain (e.g., articles in press or in preparation). Finally, the Technical

Information Center at the Naval Training Systems Center, Orlando, was searched for published technical reports within the aviation training domain.

Table 1 presents a listing of the primary sources used to obtain information for this meta-analysis. Key words that were used when locating experiments were: simulation training, training devices, simulator fidelity, training device requirements, transfer of training, training effectiveness evaluation, simulator cost effectiveness, fidelity guidance, computer simulation, simulated environment, flight training, military training, and job training.

A total of 247 journal articles, book chapters, and technical reports on training effectiveness were collected. The literature was divided into four categories: reference materials, aviation device empirical research, empirical research on other devices, and non-relevant information. The reference materials were reviewed and added to the data base if they contributed to the understanding of the empirical research. Appendix A lists experiments excluded from the meta-analysis and reasons why each was rejected. Only the experiments that involved training with a simulator and transfer to operational equipment were retained. Of those experiments, only the ones that reported the necessary statistics for meta-analysis could be included in the research integration. If an experiment lacked sufficient statistics, efforts were made to contact those who had conducted the experiment to see if they could supply the necessary data.

CODE SHEET

A code sheet was developed for use in extracting data from the collected research. This code sheet was based on the meta-model, and its purpose was to ensure that the critical information for this analysis would be collected from each report.

The initial version of the code sheet, presented in Appendix B, lists: 1) classification of task equipment; 2) training context variables; 3) the training task; 4) trainee characteristics; and 5) areas related to research design characteristics and sample population. Simulator design and fidelity level although part of the meta-model) were not included in the initial code sheet A sampling of the literature indicated that information describing the various systems which combine to make up the simulator, such as those related to motion and visual display, varied considerably from report to report. In many instances, one or more secondary sources were cited in lieu of a detailed description of the simulator. Coding of fidelity issues was delayed until more information on the simulators could be gathered.

The initial code sheet included the topic areas as major headings. An area above the headings was used for recording useful statistics and report

Table 1

Primary Sources Used to Locate Relevant Experiments

Source	Year(s)
1. Computer Search -	1982-1986
ERIC	
NTIS	
Psychological Abstracts	
2. Literature Reviews and Bibliographies	
AGARD-AR-159	1981
Ayers, Hays, Singer, & Heinicke	1984
Caro	1977
Hays & Singer	1989
Kinkade & Wheaton	1972
Martin	1981
Waag	1981
Wheaton, Rose, Fingerman, Korotkin, Holding, & Mirabella	1976

identification information. The first 25 experiments were reviewed by three individuals who coded the same experiments and discussed important characteristics found for each topic area. Through these discussions, potentially useful characteristics were identified and a more formalized code sheet was developed.

Appendix C lists items that were listed in an early version of the code sheet, but were eliminated due to lack of information in the research reports. These items include information related to research design, level of simulator fidelity, and training characteristics (e.g., proficiency based vs. blocked design, number of training trials). The final code sheet, shown in Appendix D, was developed through a series of iterations based on coders' discussions.

CODER TRAINING

Groups of two or more coders met periodically during the coding process to define coding response categories more precisely. The definitions for the coding sheet, along with any caveats, were recorded in a code book that was referred to during the coding process.

The purpose of the code book was to improve interrater reliability by providing standard information to the coders. The code book offered guidelines for selecting among coding categories, response categories, location of information, and correct calculations for numerical responses. The code book is presented in Appendix E with a description of information included for each area.

As noted in the previous section, coding the simulator configuration and fidelity level topic area was particularly problematic. Hays and Singer (1989) provided a conceptual framework that guided initial efforts in coding aspects of simulator and simulation fidelity. Pragmatic issues were explored through discussions with engineers whose expertise included simulator design and development. It was necessary to contact knowledgeable persons (e.g., primary investigators or Naval Training Systems Center project managers familiar with each device) to fill in information gaps pertaining to simulator configuration and fidelity level. When possible, several persons were contacted as a means of corroborating this information. The final response categories for this area took into consideration both conceptual and pragmatic concerns, and fused them with the additional constraint of availability of requisite information.

It should be noted that the fidelity level of a simulator was determined only for the individual subsystems that make up the simulator (e.g., visual, motion, sound). No attempt was made to give an overall fidelity rating, since it is literally impossible to assess the relative contribution of any subsystem to the simulator as a whole.

CODING PROCEDURE

At least two coders independently coded all research. The completed code sheets were discussed, and all discrepancies were resolved by consensus decision. A consensus decision process was used instead of a pooling procedure because: 1) many coding responses were discrete; 2) the consensus decision process served as a continual form of training for the coders; and 3) coders would cite information directly from the report to substantiate their coding response, thereby increasing the thoroughness of the coding task. The independently coded responses were used to calculate interrater reliability estimates, and the consensus-derived coded responses were used when performing all other analyses.

INTERRATER RELIABILITY

Among the several interrater reliability indices suggested in the literature, two are applicable to the coding procedure used in this research. In general, indices that tapped interrater agreement, as opposed to interrater consistency, were used, since the latter index allows for the possibility of having different coded responses with a demonstrated perfect interrater reliability (Tinsely & Weiss, 1975; Jones, Johnson, Butler, & Main, 1983; see also Dickinson et al., 1986). For discrete response items, Cohen's kappa (Cohen, 1968) was calculated because there were at least three response classification categories for all but two items, which were dichotomous in nature. The Cohen's kappa formula produces values ranging from - 1.0 to 1.0, with zero (0) indicating chance agreement and 1.0 indicating perfect agreement. For continuous response items, an intraclass correlation coefficient (ICC) was calculated that provides an indication of the degree to which the two coders' responses are interchangeable (Shrout & Fliess, 1979). This coefficient was used here because two coders were responsible for coding the experiments after the initial phases of formalizing the coding sheet were completed.

The reliability estimates indicated that moderate to high levels of interrator reliability were obtained using the coding procedure. For all discrete response items, the mean Cohen's kappa value was .67 and ranged from .34 to .92. When items having no variability were deleted, the mean kappa value was .69 (range was from .63 to .94). For continuous response items, the ICC (2,1) value was .95.

CALCULATION OF RESEARCH EFFECTS

There are several training outcome effect size (ES) estimates that could be used for summarizing the experiments used in this report. Glass et al. (1981) advocate use of what is commonly referred to as the d (difference) statistic, calculated by subtracting the mean performance scores of the experimental and control groups, then dividing this difference by the control group standard deviation (use of a pooled standard deviation has also been suggested). However, Hunter et al. (1982) note that this ES estimate is strongly dependent on sampling error. These researchers advocate use of either a biserial or point biserial correlation coefficient for several important reasons: first, biserial and point biserial statistics can be corrected for statistical biases from sampling error, measurement error, and restriction of range, (for both the measurement and criterion variables); second, they can be transformed into the d statistic, and are thereby readily interpretable; and third, both types of correlation coefficients can be used with multivariate analysis techniques, which have been found useful for analyzing research characteristics to identify potential moderator variables (e.g., Dickinson et al., 1986; Hunter, et al., 1982).

The point biserial correlation coefficient was chosen for use in this review because the separation of subjects into either an experimental or control group established a "true" dichotomy, a primary consideration for

determining appropriate use of the point biserial correlation coefficient (Isaac & Michael, 1978, p.126). The other criterion, that the performance measure be continuous in nature, also applies to experiments used in this review.

A detailed description of procedures used to convert one or more research statistics into a weighted mean point biserial correlation coefficient, denoted as RPB, is given in Appendix F. In general, the procedures chosen were those that would produce conservative RPBs for individual experiments, and thus the overall (population) RPB may be viewed as a conservative estimate of the flight simulator training effectiveness.

According to the Hunter et al. (1982) approach, variability involving criterion performance measures should be corrected for sampling error, unreliability, and range restriction whenever possible. In this analysis, only the correction for sampling error was used because: 1) entire classes of Undergraduate Pilot Training (UPT) graduates (or current Undergraduate Pilot trainees) were used as subjects, in many cases; and 2) usually, sampling error accounts for a majority of the spurious error relative to the other two sources (Schmitt, Gooding, Noe, & Kirsch, 1984).

With regard to transfer effectiveness, two major problems precluded attaching a dollar figure or time/training savings figure to a given RPB value. First, cumulative RPB values reported here collapse across different training programs, simulators, and tasks. Training effectiveness measures are highly dependent on training, equipment, and task variables (Orlansky and String, 1977). A second problem is related to the rapidity of technical advances in this domain. Many of the experiments included in this report were completed over ten years ago. Technology has advanced to such a degree since then that cost savings or other training effectiveness metrics related to these results may not be applicable within current simulator training programs.

Outcome measurements that were directly or indirectly based on some form of evaluator rating were considered subjective in nature. Instructor pilot (IP) ratings were the most common assessment technique for experiments reported in this review. Even seemingly objective measures, such as trials-to-proficiency, when proficiency was based on IP judgment, were classified as subjective. Only measures that were based on clearly objective indices, such as recording of instrument readings at selected points during a flight-control maneuver (Martin & Waag, 1978b), were considered objective.

Initial and final transfer trial measures were coded for all experiments that specifically reported this information. Final transfer trial information was not standard. The actual trial number used to calculate the final transfer trial RPBs ranged between the third and seventh transfer trial across experiments.

The other measure used to evaluate training effectiveness in this review was the percentage of negative research statistics. This measure is a ratio

of the number of research statistics that were negative in value (i.e., instances where the control group performance was superior to that of the experimental group), divided by the total number of research statistics used to calculate a specific RPB. Thus, for a given experiment, a percent negative research statistic measure was calculated for each valid RPB area (e.g., overall, objective only, subjective only) produced for the experiment. Although the percent negative research statistic metric is an indirect measure of training effectiveness, it does provide information about the consistency with which experimental training outcomes favored the experimental or control group.

PROCEDURE FOR DETERMINING MODERATOR VARIABLES

Experiments were coded according to both continuous and discrete response items. Since the number of experiments for each type of aircraft was small, and particularly so for helicopters, inclusion of response categories for subsequent analysis could not be guided by multivariate statistical procedures (e.g., multiple regression and factor analysis) found useful in other meta-analytic reviews (Dickinson et al., 1986; Hunter et al., 1982). Instead, individual response categories were examined using descriptive measures to assess whether there existed sufficient variability for follow-up analysis. Next, correlation coefficients were calculated between selected response categories and each dependent measure, and between each of the remaining independent variables. Finally, potential moderator variables that were identified by correlational analysis were examined further using subgroup analysis outlined by Hunter et al. (1982, p. 105; see also Dickinson et al., 1986).

The procedure for determining if a variable is a moderator using subgroup analysis was as follows. The weighted mean effect size (RPB), observed variance, error variance, and "true" variance for the total group and for individual subgroupings of___experiments were compared. Valid moderator variables produce different RPB estimates for separate subgroups when compared to each other. More importantly, the "true" variance for the individual subgroups is reduced relative to the total group. This reduction indicates that partitioning the total data set into two or more subgroups is appropriate, since these subgroups are more homogeneous in nature (i.e., show less variability) relative to the total group.

A rule-of-thumb for determining whether subgroup analysis is appropriate is given by Pearlman, Schmidt, and Hunter (1980). This rule states that 25 percent or more unexplained variance must remain after correcting for research artifacts for the total group, before it is appropriate to look for moderator variables.

THIS PAGE INTENTIONALLY LEFT BLANK

RESULTS

OVERVIEW

A total of 26 transfer of training experiments were coded; 19 for jet aircraft and seven for helicopters. Table 2 presents a brief summary of important information from each of the experiments. The last two columns describe the research statistics and the weighted mean point biserial correlation coefficients (RPBs), respectively. RPBs were calculated for five areas: 1) the overall training outcome effect (averaged across task type, transfer trial, and type of outcome measure); 2) objective measures only; 3) subjective measures only; 4) initial transfer trial; and 5) final transfer trial.

Table 3 presents a breakdown of the total number of experiments included in the meta-analysis based on aircraft type and experiment type. These breakdown variables were both conceptually and empirically based. Conceptually, since aircraft (task equipment) differ immensely in appearance and aerodynamics (jets are different from propeller-driven aircraft; and both of these aircraft are very different from vertical takeoff-and-landing aircraft), the pattern of training outcomes could be expected to differ as well. Empirically, previous reviews of flight simulation training literature have noted that findings from one type of aircraft do not necessarily generalize to other aircraft (Martin, 1981; Orlansky & String, 1977). Findings reported here support this view.

In addition to aircraft type, it was expected that different types of experiments would produce dissimilar training outcomes. Previous analyses of experiments support the contention that experiments that compare simulator training with no simulator training show different results, as a group, from those experiments that compare motion-based simulators with no motion simulators (Orlansky & String, 1977; Martin, 1981). Subgroup analysis done for this research indicate that collapsing across these two types of experiments is not meaningful.

PRELIMINARY ANALYSIS

Task Equipment

Prior to any other analysis, an analysis was performed examining whether experiments using either jets or helicopters should be treated separately or should be combined. This analysis was based on the meta-model which suggested that the actual task equipment (an input variable) may affect the training outcome. Appendix G presents the results of the subgroup analysis based on aircraft type, indicating a substantial difference between jet and helicopter experiments. The percent unexplained variance for the combined (total) group was .37, thereby exceeding the minimum .25 suggested by Pearlman et al. (1980). The subgroup analysis revealed substantial differences between the RPBs of jet and helicopter experiments. Even when results were collapsed

Table 2

Summary of Important Information for Experiments Included in the Meta-analysis

Author (yr)/Source	Population	Total N (# Grps)	Simulator/ Aircraft	Type Tasks
1. Ryan et al. (1972)/NTEC	UPT grads. [Navy]	124 (4)	2F90 (CFT)/ TA-4J	Basic instrument maneuvers-B stage of advanced jet phase training
2. Browing et al. (1973)/TAEG	UPT grads. [Navy]	26 (2)	2F69D (OFT) with 2C23A (CFT)/P-3	Tasks related to 109 item procedures/ systems checklist
3. Brictson & Burger (1976)/ NTEC	CAT I Fleet replacement pilots/varied experience [Navy]	53 (2)	(NCLT)/A-7E	Night carrier landings
4. Payne et al. (1976)/ Northrop Corp.	UPT grads. & additional pilots with varied experience	16 (2)	(LAS-WAVS)/ F-4J	8 air combat maneuvers
5. Woodruff et al. (1976)/AFHRL	UPT grads. [Air Force]	16 (2)	ASPT/T-37	Total of 4 tasks from basic to navigation
6. Browning et al. (1977)/TAEG	UPT grads. [Navy]	34 (2)	2F87F (OFT) compared to 2F69D (OFT)/ P-3	20 aircraft control maneuvers
7. Gray & Fuller (1977)/AFHRL	UPT grads. [Air Force]	24 (3)	ASPT/F-5B	10, 15, & 30 degree bomb delivery runs
8. Browning et al. (1978)/TAEG	UPT grads. with advanced flight training [Navy]	37 (2)	2F87F (OFT)/	22 tasks of varying difficulty

Table 2 (continued)

Report #	Training Length	Research Statistics	Weighted Mean Point Biserial Correlation (Mean N)
1.	15 total hours- 8hrs emergency procedures, 7hrs basic instrument/ navigation	Reported Z value, F-test using IP check flight raw scores (Table 6, p. 16)	Sim. vs. A/C only Trng. RPB(1) = .206 (63) RPB(3) = .206 (63)
2.	9 total trials lasting 8 wks. 6 in 2F69D & 3 in 2C23A	Reported P value (<.05) and group means (p.26)	Sim. vs. A/C only Trng. RPB(1) = .354 (26) RPB(3) = .354 (26)
3.	Avg. of 80 trials (ball control passes)	Reported t-tests using IP ratings & obj. perf. meas. (Table A-1, p.68)	Full vs. limited Sim. Trng. RPB(1) = .072 (144.67) RPB(2) = .065 (219.8) RPB(3) = .113 (50.75)
4.	6 trials, 1 hr per trial	Converted U values to t values (Figs. 11-13, pp. 41,42,44, & text pp. 36-38)	Sim. vs. A/C only Trng. RPB(1) = .402 (16) RPB(3) = .402 (16)
5.	Varied - mean # hours - 25.5	F-tests calculated using raw hrs to proficiency (Table 2, p.10) & IP ratings (p.12)	Sim. vs. A/C only Trng. RPB(1) = .547 (16) RPB(3) = .547 (16)
6.	6 trials, 2hrs per trial	t-tests calculated using reported mean flights to proficiency (Table 4, p. 24)	Sim. vs. A/C only Trng. RPB(1) = .606 (34) RPB(3) = .606 (34)
7.	8 trials, 1hr per trial	Reported F-tests (p. 13) and Chi square values (p. 12)	Motion vs. No-motion RPB(1) = .001 (16) RPB(2) = .017 (16) RPB(3) = .046 (16)
8.	6 trials	F-tests calculated from reported means (Table 4, p. 17)	Sim. vs. A/C only Trng. RPB(1) = .598 (37) RPB(3) = .598 (37)

Note: RPB(1)=Overall; RPB(2)=Obj. meas. only; RPB(3)=Subj. meas. only; RPB(4)=Initial transfer; RPB(5)=Final transfer;

Table 2 (continued)

Author (yr)/Source	Population	Total N (# Grps)	Simulator/ Aircraft	Type Tasks
9. & 10. Martin & Waag (1978a)/ AFHRL	UP trainees with least flight experience [Air Force]	24 (3)	ASPT/T-37	7 basic flight control maneuvers
11. & 12. Martin & Waag (1978b)/ AFHRL	UPT grads. [Air Force]	36 (3)	ASPT/T-37	8 aerobatic flight maneuvers
13. & 14. Ryan et al. (1978)/TAEG	UPT grads. [Navy]	95 (4)	2F87F (OFT)/ P-3	3 landing tasks
15. Nataupsky et al. (1979)/ AFHRL	UPT grads. [Air Force]	32 (4)	ASPT/T-37	3 basic control maneuvers
16. Reed & Reed (1979)/AFHRL	Student pilots [Air Force]	21 (3)	Air refueling director lights trainer/F-4C & KC-135	10 tasks related to air refueling
17. Martin & Cataneo (1980) /AFHRL	UP trainees (13 were AF Academy grads.)	24 (3)	ASPT/T-37	3 basic control maneuvers
18. Pierce (1983) /AFHRL	UPT grads. [Air Force]	40 (2)	ASPT/A-10	5 basic control maneuvers

Table 2 (continued)

Report #	Training Length	Research Statistics	Weighted Mean Point Biserial Correlation (Mean N)
9. & 10.	10 trials	Reported t-tests (Table E3, p.38) & calculated F-tests from IP ratings (APP. D1, pp.30-32)	Sim. vs. A/C only Trng. RPB(1)&(3) = .552 (24) Motion vs. No-Motion RPB(1)&(3) = .081 (14.7) RPB(4) = .094 (14) RPB(5) = .069 (15.4)
11. & 12.	5 trials, 5.5 total hrs.	Reported t-tests (Table D3, pp. 29-30)	Sim. vs. A/C only Trng. RPB(1) = .130 (36) RPB(2) = .023 (36) RPB(3) = .248 (36) Motion vs. No-motion RPB(1) = .101 (24) RPB(2) = .201 (24) RPB(3) = -.01 (24)
13. & 14.	6 trials	P value (<.05), (p. 12 - group C-3 vs. E)	Sim. vs. A/C only Trng. RPB(1)&(2) = .383 (29) Motion vs. No-motion RPB(1)&(2) = -.297 (50)
15.	4 trials	Reported F-tests (Table 9, p. 14)	Motion vs. No-motion RPB(1)&(3) = .138 (30) RPB(2) = .112 (30)
16.	1 hour	F-test calculated from raw IP ratings (Table 3, p. 19)	Sim. vs. A/C only Trng. RPB(1) = .141 (21) RPB(2) = .072 (21) RPB(3) = .211 (21) RPB(4) = .444 (21) RPB(5) = -.304 (21)
17.	3 trials - 1 hr per trial	Reported t-tests (Table 9, p. 18)	Sim. vs. A/C only Trng. RPB(1)&(3) = .301 (23) RPB(4) = .265 (23) RPB(5) = .329 (23)
18.	5 trials - total	Reported F-tests (Tables C-3 to C-6, pp. 30-35)	Sim. vs. A/C only Trng. RPB(1) = 1.35 (40) RPB(2) = .095 (40) RPB(3) = .148 (40)

Note: RPB(1)=Overall; RPB(2)=Obj. meas. only; RPB(3)=Subj. meas. only; RPB(4)=Initial transfer; RPB(5)=Final transfer;

Table 2 (continued)

Author (yr)/Source	Population	Total N (# Grps)	Simulator/ Aircraft	Type Tasks
19. Westra et al. (1986)/NTSC	Students entering FCLP phase of trng.	126 (2)	VTRS/T-2C	Carrier qualification landings & field carrier landing practice (FCLP)
20. & 21. Caro & Isley (1966) /HumRRO	Warrant Officer in 4 week rotary wing course [Army]	132 (4)	Whirlymite helicopter trainer/OH-23D	Basic contact flight maneuvers
22. Holman (1979)/ARI	Student pilots [Army]	59 (2)	Helicopter flight sim./ CH-47	32 control tasks, basic and advanced
23. & 24. Isley et al. (1968)/ HumRRO	Student pilots (instrument phase) [Army]	145 (3)	I-CA-I/TH-13T	4 tactical instrument flight maneuvers
25. McDaniel et al. (1983)/	UPT grads. [Navy]	26 (2)	2F64C (OFT) & 2C44 (CPT)/ SH-3H (Sea King)	9 flight control maneuvers
26. Caro et al. (1984)/NTEC	CAT I UPT grads [Navy]	22 (2)	LCCPT & 2C44 (CPT)/SH-3H	19 tasks common to both simulators

Table 2 (continued)

Report #	Training Length	Research Statistics	Weighted Mean Point Biserial Correlation (Mean N)
19.	Study variable, 20, 40, or 60 trials	Reported F-tests (Tables 10, 12, & 14, pp. 26, 40, & 43)	RPB(1) = .113 (126) RPB(2) = .162 (126) RPB(3) = .014 (126) RPB(4) = .133 (126)
20. & 21.	Study variable, either 3 or 7hrs	t-tests calculated from IP ratings & total flight time (Tables 2 & 5, pp. 42-43)	Sim. vs. A/C only Trng. RPB(1),(3),&(4)=.033(45) Limited vs. Full Trng. RPB(1),(3),&(4)=-.001(51.33)
22.	Minimum 1hr for basic tasks & 15 hrs for advanced	Reported t-tests & calculated F-tests from raw scores	Sim. vs. A/C only Trng. RPB(1)&(3) = .076 (59.5) RPB(5) = .073 (61)
23. & 24.	Study variable, either 10 or 20 hrs - total 8 weeks	t-tests calculated from reported IP ratings & error rates (Tables 6, 7, & 9, pp. 13-15)	Sim. vs. A/C only Trng. RPB(1) = -.028 (65.71) RPB(2) = -.099 (40.5) RPB(3) = -.021 (69.92) RPB(4) = .016 (89.5) Limited vs. Full Trng. RPB(1) = -.027 (56.17) RPB(2) = -.061 (39) RPB(3) = -.022 (59.6) RPB(4) = 0.0 (63)
25.	12 trials, 1 hr, 45 mins. per trial	Reported correlation coefficients (Tables 6 & 12, pp.26 & 32) & F-tests calculated from mean trials to prof. (Tables 5 & 11, pp.25 & 31)	Sim. vs. A/C only Trng. RPB(1)&(3) = .205 (25.5)
26.	6 trials, 2.5 hrs per trial	Reported t-tests (Table C-1, p.48)	Low vs. High Fidelity Simulation Training RPB(1)&(3) = .314 (20)

Note: RPB(1)=Overall; RPB(2)=Obj. meas. only; RPB(3)=Subj. meas. only;
RPB(4)=Initial transfer; RPB(5)=Final transfer;

Table 3

Breakdown of Simulator Training Experiments Based on
Aircraft Type and Experiment Type

Combined Group	Breakdown Variable	
	Aircraft type	Experiment type
		(A1) Simulator vs. Aircraft-Only Training
		$(N=10 \ / \ \overline{RPB}= .26)$
	(A) Jet Experiments	(A2) Motion vs. No Motion
	$(N=19 \ / \ \overline{RPB}= .15)$	$(N=5 \ / \ \overline{RPB}= -.05)$
		(A3) Other
		$(N=4 \ / \ \overline{RPB}= .19)$
All Experiments $(N=26 \ / \ \overline{RPB}= .13)$		
		(B1) Simulator vs. Aircraft-Only Training
	(B) Helicopter Experiments	$(N=3 \ / \ \overline{RPB}= .02)$
		(B2) Motion vs. No Motion
	$(N=7 \ / \ \overline{RPB}= .04)$	$(N=1 \ / \ \overline{RPB}= N/A)$
		(B3) Other
		$(N=3 \ / \ \overline{RPB}= .04)$

Note. "N" refers to the number of experiments at each level of breakdown variable. RPB is the weighted mean point biserial correlation coefficient computed for a specific breakdown level. N/A indicates insufficient number of experiments to compute a RPB.

across experiment type, there remained stark differences between jets and helicopters (RPBs equal .15 and .04, respectively). In Table 3, (A) versus (B) denotes this comparison. A more valid comparison, and the one used for the subgroup analysis, involved contrasting results from similar experiments for each type of aircraft. This contrast is even more pronounced (RPBs are .26 and .02 for jet and helicopter experiments; see Table 3, (A1) versus (B1)). In addition, there was a reduction of the "true" variance for jet experiments (.011) and helicopter experiments (0.00) compared to the variance of the total group (.015).

The correlational and subgroup analyses for helicopters were not conducted due to the paucity of useable experiments. The remainder of the findings from the meta-analysis are reported below. These findings pertain to jet experiments only.

Research Objectives

To test whether all experiments involving jets should be viewed together or separately, a second analysis was performed, based on the stated research goals. This analysis was conducted to assure that experiments that are dissimilar in important ways were not combined to provide meaningless results. The results of this analysis are presented in Appendix H. It was found that experiments comparing simulator and aircraft-only training appear to be substantially different from those that investigate the effectiveness of simulator motion in training (note differences in RPBs for (A1) and (A2) subgroups in Table 3). The four remaining experiments ("other" category) could not logically be combined with either of the other two types of jet experiments, and no further analyses could be performed with them. These four experiments include: 1) two that compare full simulator training versus limited training (Brictson & Burger, 1976; Pierce, 1983); 2) one that compares training using an older and supposedly lower fidelity simulator with a newer, higher fidelity simulator (Browning, Ryan, Scott & Smode, 1977); and, 3) one that compares the combined use of a cockpit familiarization trainer (CFT) and an operational flight trainer (OFT) with the OFT alone (Browning, Ryan & Scott, 1973).

In summary, initial analyses of research data demonstrate that task equipment does have an effect on training outcomes, and supports the separation of simulator training outcomes across different aircraft types. They also support the separation of training results from experiments that differ substantially in design characteristics.

Frequencies and Mean Values

Frequency data for experiments involving jets provide useful information about the research domain (see Appendix D for frequencies of individual response categories). All experiments in this analysis involving jets were reported in technical reports, and most experiments (N=10) compared simulator versus aircraft-only trained groups. Five others compared subjects trained on a simulator with the motion system turned on, with subjects trained with the

motion turned off. With the exception of two experiments, all subjects were current UPT trainees or were recent UPT graduates.

As to simulator design features (part of throughput in the meta-model, page 19), most experiments reviewed here were performed using whole task trainers with a computer generated image (CGI) system and a motion system having between one and six degrees of freedom (DOF). Field of view was reported in some experiments, with little variation. G-seats were used infrequently and use of G-suits was reported in only a few experiments. All experiments used subjective measures, tied directly or indirectly to instructor pilot ratings, although only one reported intra-rater reliability estimates. Less than a third explicitly reported information that would allow measurement of initial or final transfer performance.

SIMULATOR TRAINING: JETS

General Findings: Simulator vs No Simulator

Table 4 presents RPBs, mean percent negative research statistic values, and total number of experiments for the five result classification categories. The RPB reported in the "overall effect size" category (equal to .26) is identical to the "simulator versus aircraft-only training" category presented in Table 3 (level A1). These data indicate that experiments using objective measures reported smaller training outcomes than those using subjective measures (Table 4, (2) versus (3)). The data also show that the RPB for the initial transfer trial was noticeably greater than for the final transfer trial ((4) versus (5)), although the data should be viewed with caution due to the small number of experiments used to calculate these values. Finally, the low mean percent negative values show that the majority of the research statistics used to calculate individual RPBs were positive. This indicates a consistent training effect across the performance measures used.

In accordance with suggested guidelines for reporting results of meta-analytic reviews (Wolf, 1986, pp. 9-65), 95 percent confidence intervals (CIs) were calculated for important RPB values. These values, along with relevant statistical information, are presented in Appendix I. For Table 4, only the final transfer category incorporated a value of zero within the stated CI parameters indicating that the effect may not be very strong.

Item and Response Category Reduction. The original code sheet contained both continuous and discrete response categories for describing research characteristics. These were reduced to four continuous and six discrete response categories based on frequencies and correlational analysis. These ten research characteristics were used during subsequent analysis. Item and response category reduction is described below.

Response categories having zero frequency were eliminated. When possible, response categories were combined to allow for meaningful

Table 4

Weighted Mean Point Biserial Correlation Coefficients, Mean Percent
Negative Research Statistics, and Number of Experiments by Result
Classification Category

Simulator versus Aircraft-Only Training (JETS)

Result Classification Category	Dependent Variable		
	RPB	Mean Percent Negative Research Statistics	N
(1) Overall effect size	.26	8.1	10
(2) Objective measures only	.12	1.7	5
(3) Subjective measures only	.25	10.5	10
(4) Initial transfer trial	.19	0.0	3
(5) Final transfer trial	.03	25.0	2

Note. RPB refers to the weighted mean point biserial correlation coefficient and N refers to the total number of experiments used when calculating an individual RPB. Also, the RPB reported for classification category (1) collapses across transfer trial and measurement type.

interpretation. For example, two response categories under "subject assignment" (use of matching and the combined use of matching and random assignment) were merged into a single category to directly assess the use of matching prior to subject assignment (Table 5). All continuous items were analyzed using correlational analyses only.

There were five result classification categories for RPB measures (see Table 4, (1)-(5)). Correlational and subgroup analyses were performed using only the "overall" RPB measure (category 1), since this measure was calculated for all experiments. Furthermore, the percent negative research statistic measure was used only for correlational analysis for two reasons: 1) it is an indirect measure of training effectiveness; and 2) subgroup analysis using this measure is inappropriate. As a ratio of statistical values within a given experiment, it precludes the use of appropriate meta-analytic procedures, such as attaching weights (number of subjects) to individual experiment outcome values. Thus, calculation of a weighted mean value across experiments is not possible.

Table 5

Correlations Between Research Characteristics and Dependent Variables ,

	Dependent Variables	
Research Characteristic	$\overline{RPB}$	Percent Negative Research Statistics
(1) Use of matching prior to subject assignment	.576 (10) p=.041	.018 (10) p=.480
(2) Use of CGI visual system	-.391 (7) p=.193	.258 (7) p=.288
(3) *Total FOV of visual system	.123 (9) p=.376	-.055 (9) p=.444
(4) *DOF of motion system	.677 (10) p=.016	-.185 (10) p=.305
(5) Use of G-seat	.202 (8) p=.315	-.227 (8) p=.294
(6) Use of whole-task simulator	.593 (10) p=.035	.131 (10) p=.359
(7) Use of proficiency-based training	.639 (9) p=.032	-.331 (9) p=.192
(8) Having both objective and subjective dependent measures	-.772 (10) p=.004	.094 (10) p-.398
(9) *Number training hours	.702 (9) p=.018	-.243 (9) p-.265

Table 5 (continued)

---

(10) *Number training trials	.420	.347
	(5)	(5)
	p=.241	p=.283

* indicates variables are continuous.

NOTE. $\overline{RPB}$ is weighted mean point biserial correlation coefficient. Number in parenthesis is the number of experiments used to calculate Pearson correlation coefficient. BOLD print indicates correlation coefficient has p<.05.

Correlations. Correlations were computed between individual research characteristics and between research characteristics and the two dependent variables (RPBs and percent negative research statistics). Table 5 presents correlations between the ten research characteristics and the two primary dependent measures.

The correlation between $\overline{RPB}$ and percent negative research statistic values is negative (r= -.47, p=.087). This negative relationship was expected, since RPBs for individual experiments would tend to be higher when fewer negative research statistics were included in the effect size calculation. It follows that the correlations between individual research characteristics and the two dependent measures would take on opposite values. This pattern was not observed for three of the ten research characteristics (see Table 5, numbers (1), (6), and (10)) and is important as a criterion for determining valid moderator variables.

Intercorrelations among research characteristics from experiments included in the meta-analysis are presented in Appendix K. The nature and pattern of these may influence correlations between research characteristics and RPB measures reported in Table 5.

Examination of intercorrelations among the research characteristics is useful for understanding the relationship between these characteristics and measures of training outcomes reported here. In particular, use of objective and subjective measures when calculating training outcomes appears to be an important factor mediating the observed relationship between RPB values and several research characteristics. In addition, research variables that are indicative of the strength of training, such as number of training hours, the number of training trials, and use of proficiency-based criteria for learner advancement, were found to be positively related to RPB values. The relationship between research characteristics and training outcomes is explored further in the next section using subgroup analyses.

The six discrete research characteristics were included in the subgroup analyses involving jet experiments. The purpose of these analyses were

twofold. First, they provided statistically rigorous tests for determining which research characteristics were moderator variables; second, they provided valuable information about the magnitude of the relationship between research characteristics and training outcomes.

Four of the six discrete research characteristics had subgroupings formed by partitioning experiments into those incorporating the characteristic ("yes" grouping) and those not incorporating it ("no" grouping). The procedure for conducting subgroup analysis with these variables was identical to that presented in the overview section described earlier (see Hunter et al., 1982), with one exception. That is, a valid moderator should exhibit a reduction in the "true" variance for the "yes" group, relative to that of the whole group. The "no" group may or may not exhibit a reduced variance, since experiments falling into this category could be heterogeneous in nature. This particular procedural variation was used by Dickinson et al. (1986, p. 38) in their meta-analytic review of work performance ratings. The two research characteristics not involving a yes/no dichotomy required a reduction of variance for each of the response categories. They are the use of blocked or proficiency-based training, and the use of a whole or part-task simulator. To summarize, the criteria for determining if a research characteristic was a valid moderator were: 1) the variable produced correlations with the two dependent measures having opposite signs (see Table 5); and 2) the subgroup analysis produced a reduction in the "true" variance for individual subgroups relative to the that of the whole group.

Results of the subgroup analysis are presented in Appendix J. Two discrete research characteristics failed to meet the first criterion stated above: namely, use of matching prior to group assignment and use of a whole or part-task simulator. In addition, only one of the ten experiments reported using a G-seat, thus precluding subgroup analysis for this variable.

Input Variables

Task Requirements. Since it has been suggested that simulator training works better for some tasks than others (Orlansky & String, 1977; Semple, Hennessy, Sanders, Cross, Beith & McCauley, 1981), a grouping of experiments by task type was made. Appendix L presents results of the subgroup analysis based on three tasks: takeoffs, approaches, and landings. These tasks were chosen because relevant statistical information about each task was presented in three separate reports. Analysis based on task type was not performed because requisite task-specific information was not included in most reports.

The results of the subgroup analysis for tasks indicated a substantial improvement of training outcome (RPB) measures for these three tasks relative to that of average simulator training outcomes (RPBs equal .65, .64, and .57 for takeoffs, approaches, and landings, respectively). It is important to note that these results are based on information from three experiments and that only the approach (to landing) task realized variance reduction relative to the whole group. In addition, there was considerable variance left

unaccounted (greater than 40 percent) for both takeoff and landing tasks, thus, leaving open the question of potential moderator variables for these tasks.

Trainee Characteristics. Differences between trainees were rarely studied. Flight experience received mention in some experiments as a possible variable, but, there was insufficient information on any trainee characteristics to perform an analysis.

Throughput Variables

Simulator Design. Subgroup analysis indicated that use of a CGI visual system was not a valid moderator variable, since separating experiments according to this feature did not produce a reduction in the true variance in accordance with the prescribed criterion. Both correlational and subgroup analysis indicated that use of a CGI visual system may produce below average training outcomes (RPBs are .20 and .26 for jet experiments using a CGI visual system and for jet experiments overall, respectively). The effects of using a G-seat or G-suit were not conclusive, since only one experiment reported using a G-seat, and use of G-suits was not addressed by most experiments. Training differences based on the use of a whole or part-task simulator could not be determined, since only two experiments used a part-task simulator for training.

Table 6 presents RPB values and mean percent negative research statistics for motion experiments. Differences can be found by comparing these results with experiments investigating simulator training per se, described in the previous section (see Table 4). First, as noted before, the RPB value for motion-based experiments (-.05) was substantially different from that reported for simulator versus aircraft training experiments (.26). The negative RPB value implies that motion-based training may be detrimental to training outcomes, compared to fixed-base simulator training. This result was also reflected in percent negative research statistic values (.08 versus .44 for simulator training and motion experiments, respectively).

In summary, these data support previous research indicating that use of motion simulation for jets does not consistently produce greater training outcomes relative to simulator training without motion (Martin, 1981; Orlansky & String, 1977). Over 40 percent of research statistics comparing simulator training with and without motion favor training without motion (see Table 6).

Training Context. The type of training, proficiency or blocked, was demonstrated to have a moderating effect on the training results. Experiments that incorporated a proficiency criterion for advancing trainees produced consistent, sizable improvements in training outcomes compared to those incorporating blocked training, and compared to overall jet training outcomes (RPBs are .54, .21, and .26, respectively). This result was supported by correlational analysis.

Table 6

Weighted Mean Point Biserial Correlation Coefficients, Mean Percent
Negative Research Statistics, and Number of Experiments by Result
Classification Category

Motion versus No Motion Experiments (JETS)

Level of Research Characteristic/ Result Classification Category	Dependent Variable		
	$\overline{RPB}$	Mean Percent Negative Research Statistics	N
(1) Overall effect size	-.05	44.0	5
(2) Objective measures only	.11	32.0	3
(3) Subjective measures only	-.04	34.6	5
(4) Initial transfer trial	.12	16.5	2
(5) Final transfer trial	N/A	N/A	-

Note. RPB refers to weighted mean point biserial correlation coefficient
and N refers to the total number of experiments used when calculating an
individual RPB. Also, the RPB reported for classification category (1)
collapses across transfer trial and measurement type.

Experiments incorporating both objective and subjective evaluation
reported training outcomes of lower magnitude than those using only subjective
measures. This relationship appeared to influence the observed correlation
between RPB values and several other research characteristics.

It was also expected that experiments having greater numbers of training
hours would produce higher training outcomes. The correlation for training
hours with RPB measures ($r = .70$) was positive and significant. It should be
noted that total training hours and total training trials were calculated by
collapsing across all task boundaries within a given experiment. A more
meaningful measure would have been to calculate the average number of training
hours or trials per task, but information needed to calculate per-task
averages was not given in most experiments.

DISCUSSION AND CONCLUSIONS

OVERVIEW

Review of the aviation training effectiveness research clearly shows that relatively few of the potential moderating variables have been incorporated into flight simulation experiments. Some of these variables, especially those involving experiment quality, have a significant influence on how experimental results are interpreted and may affect the magnitude of training outcomes (see Appendix C). Of those variables that have been incorporated into this research, this meta-analysis found several that have a clear moderating effect on training transfer.

Sizable differences in the effectiveness of simulation training were found between jets and helicopters. This section focuses on each of the topic areas of the meta-model (Figure 1) used as the framework for developing the code sheet. Important issues within each area are addressed, and results from previous reviews are presented. This discussion also provides a research agenda which can be used both to guide future research efforts in the flight simulation training area, and to suggest ways of documenting efforts to maximize their value for future meta-analytic reviews.

INPUT VARIABLES

Task Equipment

Not all flight simulators and training systems incorporating these simulators are the same. That this dissimilarity extends to the training effectiveness of the simulators is supported by the results presented here. In particular, there are dramatic differences in both the magnitude and pattern of training outcomes for jet and helicopter simulator training systems. For jets, simulator training outcomes have been consistent and positive. That is, comparisons between pilots trained in the aircraft only, and those trained on a simulator and the aircraft, consistently favored the latter group (RPB= .26). This pattern is true across a variety of task boundaries. For helicopters, the accumulated difference between simulator and aircraft only trained pilots was quite small (RPB= .02). Over 40 percent of the experimental comparisons favored the aircraft-only trained group (compared to eight percent for jets). Experiments directly assessing use of simulator motion indicated motion cuing did not improve training for jets (see Table 3, A2). Results from the sole experiment included in this review assessing the effects of motion cuing for helicopters (McDaniel et al., 1983) indicated that certain helicopter tasks may benefit from motion cues.

The observed differences in experimental results for these two aircraft types necessitated separate analysis and reporting of results of jet and helicopter experiments. The limited number of helicopter experiments that could be included in this review precluded any in-depth analysis aimed at identifying moderator variables in this area. For these reasons, the following discussion is confined to training involving jet aircraft, except

where otherwise stated. The subject population targeted for this review was novice jet pilots; specifically, recent Undergraduate Pilot Training (UPT) graduates or current trainees. Thus, results of this meta-analysis are not generalizable to transition pilots with prior jet experience or pilots having extensive prior simulator experience.

Task Requirements

Previous reviews (Orlansky & String, 1977; Semple et al., 1981) suggested that certain "basic" aircraft control tasks (e.g., approach and landing) appeared to transfer much better than more complex tasks (e.g., formation flight maneuvers). Since no existing task taxonomy was appropriate for the tasks in this domain, an attempt was made to develop task categories based on difficulty ratings assigned to tasks by novice and experienced pilots. Appendix M presents the rating form used to collect task difficulty information for jet aircraft. This effort was only partially successful. There are several factors that must be considered before task difficulty information can be a useful metric. For example, task difficulty is relative, so tasks trained early in the training program (e.g., descending turns) may seem difficult until one attempts more complex tasks during a later training phase (e.g., carrier landings). Thus, only a few tasks were rated in a consistent manner along a continuum ranging from "low" to "high" difficulty.

Individual research reports were examined for inclusion of the same task or set of tasks, with training outcome data reported for individual tasks, to allow for calculation of a cumulative RPB value for each task. Three tasks were found that met these criteria: normal takeoffs, approaches, and landings (excluding carrier landings). Cumulative RPB values were over two times greater for these tasks (RPBs= .65, .64, and .57, respectively) relative to the overall value for jet training (RPB= .26).

These results underscore the need for reporting task-specific training outcome information in future research efforts. Without the inclusion of such data, as well as detailed information about simulator fidelity and configuration parameters, future meta-analytic reviews will not be able to quantify performance outcome tradeoffs for varying fidelity levels of specified simulator subsystems.

Trainee Characteristics

Student pilots bring with them into the learning environment different aptitudes, abilities, and prior experiences. These factors can influence the amount and rate of knowledge acquisition, retention, and transfer of training. Taken together, these factors comprise what is commonly referred to as individual differences.

In their review of individual differences within military training environments, Hogan, Arneson, and Salas (1987) cite evidence suggesting that individual difference factors may account for a portion of the variance associated with simulator training outcomes. For example, Federico (1982)

presented findings indicating that even when training programs incorporate mastery-level criteria for advancing or terminating training, differences between subjects' performance are still noticeable. Flammer (1976) reported that mastery training did not reduce individual differences in learning time within a given mastery unit (see also Arlin, 1984).

Motivation is an individual factor thought to have considerable influence for both initial skill acquisition and for subsequent transfer to the operational environment (AGARD Report, 1980). The authors of the 1980 AGARD report considered understanding and resolving motivational issues to be the key to maximizing training outcomes. Since motivation plays an important role in current theories of learning (e.g., Bandura & Walters, 1963; Gagne', 1985; Skinner, 1953) and instructional development (e.g., Dick & Carey, 1978), it may be considered to influence all phases of simulator-based training, from device design to evaluation of training performance. Motivational issues include both the acceptance of the simulator as a valid training device, and the design of training that involves and challenges the student.

There is evidence that students may lose motivation after prolonged simulator training simply because they would rather begin training in the aircraft (Pohlmann & Reed, 1978, p. 8). Despite the possible influence of motivation, the flight training research included very little pertinent information about it. Formal assessment of instructor and student acceptance of a given simulator was rarely attempted (see e.g., Reed & Reed, 1979), although anecdotal information was given in a few reports.

One other attempt was made to investigate the effects of motivation within training. Performance feedback, in the form of knowledge of results (KOR), has been shown to have motivational properties (Kulhavey, 1977; Kulhavey, White, Topp, Chan, & Adams, 1985). It was thought that differences in how KOR was given, either instructor generated, device generated, or a combination of both, would influence performance outcomes. Unfortunately, only three reports clearly specified how performance feedback was given (Gray & Fuller, 1977; Payne et al., 1976; Westra, Lintern & Wightman, 1986). This number was insufficient for meaningful analysis.

In general, researchers within the flight simulation training area have treated individual differences between learners as a potential source of error variance. Seven of the ten most common types of jet experiments included in this review used a matching procedure to equate subjects prior to group assignment. The matching variables used most frequently were overall UPT scores and previous number of flight hours.

Use of matching prior to subject assignment correlated positively with both RPB and percent negative research statistic values (r's = .58 and .02, respectively). However, since the pattern of results of the correlational analysis for this variable did not conform to that prescribed for selecting variables for additional (subgroup) analysis, it was eliminated as a potential moderator variable, and no further analysis was done.

THROUGHPUT VARIABLES

Simulator Design

One of the goals of this analysis was to identify simulator fidelity configuration parameters that optimize training outcomes. To accomplish this goal, simulators were viewed in terms of individual subsystems, and when possible, an attempt was made to evaluate separate components or design features within a given subsystem. For example, the motion and force cuing subsystem was separated into use of G-seat and DOF of the platform motion apparatus. Similarly, evaluation of the visual image generation subsystem involved separate analysis of FOV parameters as well as use of CGI technology. At a more global level, analysis was performed based on whether the simulator was considered a whole or part-task device. Lack of variability in the reported use of other design features precluded any attempts at analysis. These included use of G-suit force cuing, sound simulation, "stick shaker" system, instructional and environmental features, as well as the type of procedure used to validate the flight control characteristics.

The gathering of information for analysis on fidelity of simulation was hampered by two factors: the deficiency of detailed reporting of simulator configuration parameters, and the lack of a validated taxonomy of flight tasks. The latter compelled the assessment of fidelity information on a task-by-task basis. Although several task taxonomies have been developed (see e.g., Wheaton et al., 1976; Fleishman & Quaintance, 1984), none were found to be appropriate for use in aviation tasks.

The findings reported here are only an initial step toward fulfilling the goal of extracting empirically-based design guidance principles, because detailed information on the training device used is not routinely reported in research reports. As a result, the level of analysis possible from available information may be too global to be of immediate use by engineers and other simulator design specialists.

Visual Simulation. The only two variables used to evaluate visual imaging systems were total FOV of the system and use of CGI technology. A more thorough evaluation such as determination of configuration requirements would be more helpful but is not possible from the information available. Neither total FOV nor the use of CGI technology was found to have a moderating effect on training outcomes. This is consistent with findings reported elsewhere (Sample et al., 1981; Woodruff, Smith, Fuller, & Weyer, 1976).

This result should not be taken to indicate that visual imaging is an unimportant factor in simulator training. Recent surveys of engineers and other training specialists, to determine human perception and performance information needed when making design decisions, found visual and motion simulation areas were the two most frequently stated areas of need (Klein & Brezovic, 1987; Rouse, 1983). Additionally, the AGARD (1980) report

46

concluded that "... With few exceptions, the overwhelming finding has been that visual tasks learned in the simulator show positive transfer to the aircraft" (p. 9). Finally, visual imaging technology is far superior today to that used to produce the visual systems of simulators used in this review. For these reasons, continued evaluation of current training systems that incorporate this technology is warranted.

Motion and Force Simulation. Considerable interest and attention has been placed on the utility of simulator motion cuing for facilitating skill acquisition and transfer. In general, results of this meta-analysis support the previous reviews which indicate motion cuing adds little to the training environment (Martin, 1981; Hays & Singer, 1989; Orlansky & String, 1977). The cumulative effect size value across the five motion versus no-motion experiments included in the meta-analysis was negative in value (RPB= -.05), indicating that motion may detract from training, at least for some tasks.

These results are inconsistent with findings from a recent review of the flight simulation evaluation literature by Pfeiffer and Horey (1987). There are several obvious differences between the Pfeiffer and Horey (1987) review and this review that help to explain these contradictory findings. First, Pfieffer and Horey used as their research effect size metric, a transfer effectiveness ratio (see also Hays & Singer, 1989, pp. 133-134). This measure is highly dependent on the length of training on the simulator. Second, although their approach was described as "meta-analysis" (p. 15), they did not incorporate commonly accepted meta-analytic methodology, such as weighting individual research outcomes by their corresponding sample size, or providing a detailed explanation describing their rationale for decisions/procedures used. Finally, the comparative analysis upon which they concluded the superiority of simulator training with motion cuing involved use of a t-test (p. 39). This statistical procedure is not appropriate where the underlying means and standard deviations were derived by collapsing across experiments (not subjects), and thereby calls into question the nature of the distribution upon which the critical value of the statistic is based.

Evidence indicating that motion cuing adds little, or nothing, to the jet simulator training environment cannot be considered definitive. There are two important reasons for questioning these results. First, calibration of critical motion cuing system parameters (e.g., control input response times, leg extension acceleration rates) was rarely attempted. Only one motion-related experiment included in this review reported results of calibration tests prior to experimentation (McDaniel et al., 1983). Since a similar assessment was not done during or after the experiment, the possibility of software or hardware failure during the course of the research is a cogent argument for training outcomes on certain tasks favoring the no-motion trained group. Incorporating appropriate methodological procedures, such as periodic calibration checks, is crucial for producing unequivocal results in this area.

A second reason for questioning results of motion versus no-motion experiments is due to the inclusion of several training tasks in each

47

experiment. It has been argued that motion effects vary from task to task depending on the primacy of motion cues for performing critical aspects of the task. Since reports often collapse across task boundaries when making between-group comparisons, possible specific effects from motion cuing may be inadvertently masked. Generally, reports do not distinguish between the kinds of motion provided.

Another issue when considering motion was addressed by Gundry (1976) who distinguished between maneuver motion and disturbance motion, the former resulting from aircraft control inputs of the pilot and the latter from environmental conditions, such as weather turbulence or mechanical malfunction. Gundry reasoned that, whereas providing motion cues related to disturbance would benefit simulator training, incorporating maneuver motion cues would not (see also Martin, 1981; DeBerg, McFarland, & Showalter, 1976). Information on the type of motion used in experiments was not reported, so the difference could not be evaluated.

In contrast to motion experiments involving jets, a similar experiment using helicopters (McDaniel et al., 1983) produced a positive overall training outcome (RPB= .21). This result must be tempered by the fact that information from a single experiment was used to derive the cumulated RPB metric and methodological problems cited above apply to this experiment. In this regard, a close examination of experimental outcomes from the McDaniel et al., (1983) experiment indicates noticeable differences in the direction of training outcomes for certain tasks. Specifically, positive training outcomes (i.e., instances where the motion group outperformed the no-motion group) were realized on three tasks: Aircraft Stablization Equipment (ASE) off, free-stream recovery, and coupled hover (RPBs= .19, .37, and .45, respectively). For all other tasks, including takeoffs, approaches, and landings, motion cuing was associated with negative training outcomes. This pattern of results indicates that motion cuing may aid only certain training tasks. Results from additional experiments of this kind must be added to these before conclusions concerning task-specific motion effects can be made with any degree of confidence.

Training Context

Some previous reviews of the flight simulation area have stressed that a systems approach be taken when evaluating simulation training (AGARD Report, 1980; Hays & Singer, 1989; Rose, Wheaton & Yates, 1985; Semple et al., 1981; Wheaton et al., 1976). According to the systems approach, the simulator is never a stand-alone item being evaluated, but must be considered along with other relevant features of the training milieu, such as curricula, scheduling, staffing, and the use of specific training procedures. The AGARD (1980) report concluded that, "...how the device is used can influence its effectiveness to an equal or greater extent" relative to that expected by appropriately matching the simulator to task parameters (p. 9). Unfortunately, the accumulated research yielded little data on this issue. This meta-analysis provided information on only two areas within the training context: training type and performance measurement.

Training Type. Use of training procedures that accommodated individual learner needs, such as those associated with proficiency-based training, were found to be more effective than procedures which allocated a fixed amount of training (RPBs were .54 and .21, respectively). The latter training is commonly referred to as blocked, lock-step, or fixed time/trials training. The results of both correlational and subgroup analysis reported here clearly indicate that proficiency training resulted in greater training transfer to the operational environment than transfer from blocked training.

Performance Measurement. The majority of research has used instructor ratings to measure transfer performance. Despite wide use of these ratings, problems associated with their use were mentioned in several reports; no inter rater reliability estimates were reported, and intra-rater reliability estimates were reported in only one experiment (Westra, Lintern, Sheppard, Thomley, Mauk, Wightman, & Chambers 1986, p. 47). The widespread occurrence of omitting inter- and intra-rater reliability information is problematic, since true differences in performance cannot be inferred unless the measures used to rate the performance are reliable (Cook and Campbell, 1979).

Objective measures of differences in pilot performance due to training variables were consistently lower than subjective measures. This finding is opposite to that reported by Sample et al. (1981, pp. 31-32). This discrepancy is due to differences in how the two reports discriminated between subjective and objective measures, and to the inclusion of different research reports in the two analyses. Trials-to-proficiency measures, when proficiency is based on instructor ratings, are considered subjective measures in this report. In their review, Sample et al. (1981) identified trials-to-proficiency data reported by Browning, Ryan, and Scott (1978) as objective measures (see Browning et al., 1981, Table 4, p. 17). There is also little overlap in the experiments included in the two reviews. Of the five experiments included in the Sample et al., (1981) review, three were not incorporated into this meta-analytic review (see Appendix A).

THIS PAGE INTENTIONALLY LEFT BLANK

AGENDA FOR FUTURE RESEARCH IN THE FLIGHT SIMULATOR TRAINING AREA

OVERVIEW

A quantitative literature review should provide a summary of the empirical findings and knowledge gaps where future research efforts should focus. It is apparent from this meta-analytic review that many important flight simulator training factors have yet to be addressed in a systematic fashion. The research agenda that follows is separated into the topic areas used as a framework for this report, input variables and throughput variables.

INPUT VARIABLES

Task Equipment

This meta-analysis has shown that there are major differences between jet and helicopter training effects. Additional research is needed to further specify these differences and determine the training methods that will provide maximum effectiveness for each type of simulator.

Task Requirements

Subgroup analysis reported here indicates that, simulator training does not provide equal benefit to all aviation tasks. Detailed descriptions of skills needed to perform tasks within a given training program are available (see e.g., Payne et al., 1976), yet these descriptions have not led to a valid taxonomy for grouping aviation tasks.

As noted previously, research programs in this area would benefit greatly if task categories (taxonomies) could be developed and validated. This would allow generalization of results from single tasks to task groupings. The search for a valid taxonomy for aviation tasks is critical for avoiding costly duplication of future research efforts. There appear to be several reasons why aviation tasks are not readily classified into well defined groups. One reason is that activities and skills needed to correctly perform the various tasks differ considerably. In many instances, psychomotor performance is required in addition to cognitive decision-making skills. Another reason is that as the student pilot acquires more flight hours and masters successive tasks, his reliance on, and use of, various informational sources may shift.

Fleishman and Quaintance (1984) discuss a number of "...descriptive schemes using behavior requirements as a basis for the classification of human task performance" (p. 127). While several of these schemes appear suitable for classifying aviation tasks, much work needs to be done before they can be applied to aviation tasks with any degree of confidence. Validation of any task classification system depends on availability of detailed performance outcome information for individual tasks. For this reason, future simulator training research should provide detailed training outcome information for

51

individual tasks. This information will allow researchers to apply appropriate statistical procedures (e.g., multivariate analysis) in order to empirically validate task clusters.

The process required to validate an aviation task taxonomy may take an extended period. Research in other areas may offer short-term payoffs in terms of empirically-derived training guidance. One area needing investigation involves determining the simulator instructional features that will improve training outcomes for specific aviation tasks.

Trainee Characteristics

Within the flight simulation training domain, trainee characteristics make up what are considered individual differences and are usually viewed as a source of measurement error. Within experiments used in this review, most equated subjects via matching variables (UPT scores or number of years flying) prior to assigning them to either the experimental or control group.

In their review of individual differences in military training environments, Hogan et al. (1987) cite evidence indicating that individual learners bring with them into the training environment cognitive and non-cognitive factors that influence training outcomes. Training programs may be customized to match individual learners in such areas as learning styles, cognitive strategies, and sensory modalities (Goodman, 1978). The underlying assumption for designing customized training programs is that individual learners vary in their approach to understanding and remembering new information. Since, for a given individual, preferred methods of learning are thought to be linked to the learner's interests, abilities, aptitudes, and motivations, training programs may facilitate or inhibit the learning process. Thus, one area that needs attention is the development of useful methods for determining a learner's cognitive and non-cognitive capabilities. Hogan et al. (1986) review several measurement batteries useful for determining a person's learning style or cognitive abilities (see also Su, 1984). Non-cognitive factors, such as personality, affective adjustment, or physical ability, can also be used to predict training success and may be used in addition to cognitive ability measures to enhance their predictive properties (Hogan et al., 1986).

Individuals differ in the level of motivation they bring to training and also in how well the training program can motivate them. One necessary area of research is how to promote acceptance of the simulator for training.

A second area of motivation investigation involves performance feedback in the form of knowledge of results (KOR). KOR may be generated by several sources within the simulator training environment, including the device (e.g., hardcopy printout of flight maneuver elements) and the instructor (e.g., verbal debrief). Future research projects should investigate the relative effects on training outcomes for each of these sources of KOR or the combined use of these sources. In addition, timing and amount of information inherent in the KOR have been found to influence performance in the

psychological and educational training literature (see Kulhavey, 1977; Anderson, Kulhavey, & Andre, 1972) and should also be investigated. Even if KOR is not manipulated within the experiment, detailed reporting of this information in future research projects will allow subsequent meta-analytic reviews of this area to extract useful guidelines for obtaining optimal training outcomes based on this variable.

THROUGHPUT VARIABLES

Simulator Design

This review did not attempt a fine-grained analysis using simulator configuration and fidelity levels. The primary reason for this was the lack of detailed descriptions of the simulator configuration parameters in use during experimentation. Results of experiments assessing the utility of motion cuing for both jets and helicopters were questioned because they lacked appropriate methodological controls, such as periodic calibration checks of the motion cuing (hardware/software) components.

These limitations suggest that several areas are in need of further research. In all cases, close attention must be paid to experimental methodology to insure that the results are free from potential competing explanations concerning the source of the observed experimental effects, and reporting in detail about the simulators used in research must be encouraged.

Visual Simulation. Technological advances have made experimental results from early visual simulation virtually obsolete. Research must be advanced, particularly in determining cue requirements for low-level flight.

Motion and Force Simulation. Interest remains high in how motion and cuing affect training. Methodological considerations are especially pertinent for accurately assessing the effects of motion cuing on training outcomes. Future research in this area should address the issue of task-specific motion effects. Detailed reporting of results for individual tasks within a given experiment will provide critical information for determining what task or sets of tasks benefit from motion/force cues. In addition, this information may also be used to extrapolate to certain emergency situations which cannot be trained in the aircraft for safety reasons.

Training Context

Factors within the training domain may provide the highest payoffs for improving training outcomes. Topic areas that are in need of investigation include: training type and performance measurement.

Training Type. The general finding reported here is that programs incorporating a proficiency criterion during training are associated with training outcomes approximately twice as large as those using blocked training procedures. Given the nature of military training, use of proficiency criteria during training may not be feasible in all instances. There are

training techniques that may partially substitute for proficiency basic training. These techniques can be applied within a blocked training program and may boost training outcomes.

For example, Bailey, et al., (1980) reported the use of a backward chaining procedure to be quite effective when training a 30-degree dive-bomb maneuver. This procedure involved breaking the maneuver into several steps, such as final approach, roll-in, base leg, and downwind leg. Training then proceeded in reverse order through the steps, thus giving the student ample practice on what was considered the most critical part of the task (i.e., the final task segment). This procedure made use of an instructional feature typically found on most full and many part-task simulators (i.e., initialization). Appropriate use of this procedure would require the instructors to learn to use relevant instructional features in order to implement the backward chaining procedure. In this regard, at least one report has presented evidence indicating instructional features incorporated within the simulator are rarely used (Gray, Chun, Warner, & Eubanks, 1981; see also Tracey, 1984). This suggests that instructional features may need to be accompanied by an embedded training program that demonstrates the application of relevant learning principles and procedures for each available instructional feature.

In general, the challenge for researchers and training developers is to devise training programs for instructors that will enhance training outcomes for blocked training programs to a level equal to programs using proficiency-based criteria. A similar challenge was given by Bloom (1984) to training developers in the psychological and educational domains; that is, to devise group training programs that will equal training outcomes expected when training is on a one-on-one basis.

Performance Measurement. A major problem with the research in this area is the almost complete reliance on subjective instructor/pilot ratings. Toward the goal of establishing improved subjective performance measures, the need to document inter-rater reliability information in experiments is required. Influential and far reaching decisions are being made based on the effectiveness of simulator training, compared to similar training in the aircraft (see e.g., Orlansky & String, 1977). Given that the metric of training effectiveness is regularly based on IP ratings, it is imperative that these measures be reliable when used in an experiment, or at the very least, that the unreliability of these measures be factored into the decision process.

SUMMARY

For this review, issues within the flight simulator training domain were separated into two major areas: input variables (task equipment, task requirements, and trainee characteristics) and throughput variables (simulator design and training context). These areas coincide with components of the meta-model depicted in Figure 1, which in turn were derived from previous reviews of the aviation training domain. A primary goal of this review was to identify variables that moderate the magnitude of simulator training outcomes, including specific design/fidelity features, from results of transfer-oftraining (TOT) experiments in this area. Experiments were included if they reported sufficiently detailed information to allow analysis involving appropriate meta-analytic techniques. A second goal of this review was to provide an agenda for future research to fill information gaps derived from results of the meta-analysis. Finally, guidelines describing information that needs to be reported in future research publications were generated to aid researchers in this area. These guidelines are presented in Appendix N, and will help to ensure that results from future experiments can be used in subsequent meta-analytic reviews.

Lack of detailed reporting of information concerning training methods, simulator configuration, fidelity levels, and training tasks hampered detailed analysis in these areas. Insufficient statistical information resulted in the exclusion of a number of experiments.

The major findings of the meta-analysis are as follows:

Task Equipment

The outcomes of the experiments involving the training of jet pilots were different from those involving the training of helicopter pilots. Results differed in both size and pattern of training outcomes. Jet experiments consistently found simulator training combined with aircraft training to be better than training in the aircraft alone. The findings from similar helicopter experiments were less consistent, and only slightly favored simulator training combined with aircraft training over aircraft training alone.

An insufficient number of helicopter experiments (total N=7) precluded any in-depth analysis involving this type of aircraft. Therefore the results of the meta-analysis are specific to jet aircraft training involving recent Undergraduate Pilot Training (UPT) graduates or current trainees with little or no experience in a simulator or in a jet aircraft.

For jets, the overall training effect for all tasks trained was positive and robust. Over 90 percent of the experimental comparisons favored the simulator and aircraft trained group over the aircraft-only trained group. On the average, subjective performance measures (e.g., instructor ratings) were more sensitive to training effects, and produced greater results than those obtained with objective measures (e.g., instrument readings). As

training for both groups progressed and reached the point where it was conducted solely in the aircraft, differences between the groups diminished.

Task Requirements

Certain tasks were more effectively trained in the simulator than others. For jets, when simulators were used for the training of takeoff, approach (to landing), and landing (excluding carrier landings) tasks, the training effects were greater than they were for the combination of all tasks.

Trainee Characteristics

Only two trainee characteristics were identified as likely to have an effect on training results, flight experience and UPT grades. These differences in trainees were rarely studied. When there was concern that these differences might affect training in any single experiment, an effort was made to compose each of the trainee groups with equal amounts of experience or similar grades.

Simulator Design

For jet training, motion cuing was found to add nothing to the simulator training effectiveness, and in some cases, may have taken away from the training value of the simulator. However, this finding may not be truly representative of the effectiveness of motion-based training since: 1) there was a lack of periodic calibration of the motion cuing systems; and 2) the results were based on all tasks combined. The positive effects of motion for any one task may have been masked by the negative effects of motion for another task.

Training Context

The average effectiveness for training programs where trainees were allowed to progress based on a demonstrated proficiency was greater than for training programs where all trainees proceeded at the same pace. Information on other aspects of the training context, such as the use of instructional features and the provision of feedback was seldom reported and could not, therefore, be analyzed.

COORDINATION

This effort had its genesis several years ago at the Army Research Institute for the Behavioral and Social Sciences (ARI) with the development and publication of <u>An annotated bibliography of abstracts on the use of simulators in technical training</u> (Ayers et al., 1984). Continued coordination with ARI has provided data and advice during the course of the effort. Two persons at ARI have been especially helpful:

> Mr. Charles Gainer
> Chief, ARI Aviation R&D Activity
> ATTN: PERI-IR
> FT Rucker, AL 36362-5354
> AV 558-4404, (205) 255-4404
>
> Dr. Michael J. Singer
> ARI Field Unit - Orlando
> 12350 Research Parkway
> Orlando, FL 32826-3224
> AV 960-4387, (407) 380-4387

Coordination and assistance were also provided by numerous engineers, researchers, and other individuals at the NAVTRASYSCEN.

THIS PAGE INTENTIONALLY LEFT BLANK

.REFERENCES

Adams, J. (1979). On the evaluation of training devices. *Human Factors*, *21*, 711-720.

AGARD (1980). *Fidelity of simulation for pilot training* (AGARD-AR-159). NATO: Advisory Group for Aerospace Research and Development.

Alberry, W., and Gum, D. (1978). *Motion and force cueing requirements and techniques for advanced tactical aircraft simulation*. (AFHRL Tech. Rep. TR-78-73). Wright Patterson Air Force Base, TX: Air Force Human Resources Laboratory.

Allen, J.A. and Hays, R.T., Buffardi, L.C. (1986). Maintenance training simulator fidelity and individual differences in transfer of training. *Human Factors*, *28*(5), 497-509.

Anderson, R.C., Kulhavey, R.W., and Andre, T. (1972). Conditions under which feedback facilitates learning from programmed lessons. *Journal of Educational Psychology*, *63*, 186-188.

Arlin, M. (1984). Time, equality, and mastery learning. *Review of Educational Research*, *54*, 65-86.

Ayers, A., Hays, R., Singer, M. and Heinicke, M. (1984). *An annotated bibliography of abstracts on the use of simulators in technical training*. (ARI Res. Prod. 84-21). Alexandria, VA: U.S. Army Research Institute.

Bailey, J., Hughes, R. and Jones, W. (1980). *Application of backward chaining to air to surface weapons delivery training*. (AFHRL Tech. Rep. TR-79-63). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Bandura, A., and Walters, R.H. (1963). *Social Learning and Personality Development*. New York: Holt, Rinehart, & Winston.

Bangert-Drowns, R.L. (1986). Review of developments in meta analytic methods. *Psychological Bulletin*, *99*, 388-399.

Barrett, G.V., Cabe, P.A., Thornton, C.L., and Kerber, H.E. (1969). Evaluation of a motion simulator not requiring cockpit motion. *Human Factors*, *11*(3), 239-244.

Baum, D.R., Reidel, S., Hays, R.T. and Mirabella, A. (1982). *Training effectiveness as a function of training device fidelity* (ARI Tech. Rep. 593). Alexandria, VA: U.S. Army Research Institute.

Besco, R.O. (1961). The effects of cockpit vertical accelerations on a simple piloted tracking task. *Human Factors*, *3*(4), 229-236.

Bickley, W. and Bynum, J. (1980). Training device effectiveness: Formulation and evaluation of a methodology (ARI Res. Rep. 1291). Alexandria, VA: U.S. Army Research Institute.

Biersner, R. J. (1976). Observations on the use and evaluation of ECII-LP simulators for aviation training (Rep. 2-76). Pensacola, FL: Chief of Naval Education and Training Support.

Billings, C., Gerke, R., and Wick, R. (1975). Comparisons of pilot performance in simulated and actual flight. Aviation, Space and Environmental Medicine, 304-308.

Bloom, B.S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 13, 4-16.

Branson, R.K., Rayner, G.T., Cox, J.L., Furman, J.P., King, F.J., and Hannum, W.H. (1975). Interservice Procedures for Instructional System Development (5 Vols.). TRADOC Pam 350-30 and NAVEDTRA 106A. Ft. Monroe, VA: U.S. Army Training and Doctrine Command.

Brictson, C.A. and Breidenbach, S.T. (1981). A7E Training effectiveness evaluation: The use of the night carrier landing trainer (NCLT), Device 2F103, for night carrier Qualification CO) training (Naval Training Equipment Center, Contract N61339-77-D-0028). La Jolla, CA: Dunlap and Associates, Inc.

Brictson, C., and Burger, W. (1976). Transfer of training effectiveness: A73 night carrier landing trainer (NCLT) device 2F103. (NAVTRAEQUIPCEN Tech. Rep. 74-C-0079-1). Orlando, FL: Naval Training Equipment Center.

Brown, A.L., and Campione, J.C. (1984). Three faces of transfer: Implications for early competence, individual differences, and instruction. In M. Lamb, A. Brown, & B. Rogoff (Eds.), Advances in Developmental Psychology (Vol. 3, pp. 143-192). Hillsdale, NJ: Erlbaum.

Browning, R.F., McDaniel, W.C., and Scott, P.G. (1981). Preparation and design for a training effectiveness evaluation of device 2F64C for replacement pilot training (TAEG Tech. Rep. 108) Orlando, FL: Training Analysis and Effectiveness Group.

Browning, R.F., McDaniel, W.C. and Scott, P.G. (1982). An assessment of training effectiveness of device 2F64C for training helicopter replacement pilots (TAEG Tech. Rep. 127). Orlando, FL: Training Analysis and Effectiveness Group.

Browning, R.F., Ryan, L.E., and Scott, P.G. (1978). Utilization of device 2F87F OFT to achieve flight hour reductions in P-3 fleet replacement pilot training (TAEG Tech. Rep. 54). Orlando, FL: Training Analysis and Evaluation Group.

Browning, R., Ryan, L., Scott, P., and Smode, A. (1977) Training effectiveness evaluation of device 2F87F, P-3C operational flight trainer (TAEG Rep. 42). Orlando, FL: Naval Training Equipment Center.

Bryan, G.L., and Regan, J.J. (1972). Training system design. In H.P. Van Cott and R.G. Kinkade (Ed.s) Human engineering guide to equipment design. Washington, D.C.: American Institutes for Research.

Bullock, R., and Svyantek, D. (1985). Analyzing meta-analysis: Potential problems, an unsuccessful replication, and evaluation criteria. Journal of Applied Psychology, 70, 108-115.

Burger, W., and Briccson, C. (1976). A7E transfer of training effectiveness: Device 2C15A CPT and device 2F84B OFT/WST (NAVTRAEQUIPCEN Tech. Rep. 74-C-0079-2). Orlando, FL: Naval Training Equipment Center.

Bynum, J.A. (1978). Evaluation of the singer night visual system computer-generated image display attached to the UH-1 flight simulator (ARI-RR-1199). Ft. Rucker, AL.: (U.S. Army Aviation Center) Director of Training Developments. (AD A060566).

Caro, P.W. (1970). Equipment-device task commonality analysis and transfer of training (HumRRO Professional Paper 70-7). Alexandria, VA.: Human Resources Research Organization.

Caro, P.W. (1972) Transfer of instrument training and the synthetic flight training system (Professional Paper 7-72). Alexandria, VA: Human Resources Research Organization.

Caro, P.W. (1977). Some factors influencing Air Force simulator training effectiveness (HumRRO Tech. Rep. 77-2). Bolling Air Force Base, D.C.: Human Resources Research Organization.

Caro, P.W. (1979). The relation between flight simulator motion and training requirements. Human Factors, 21, 493-501.

Caro, P.W., Corley, W.E., Spears, W.D. and Blaiwes, A.S. (1984). Training effectiveness evaluation and utilization demonstration of a low-cost cockpit procedure trainer (NAVTRAEQUIPCEN Tech. Rep. 78-C-0113-3). Orlando, FL: Naval Training Equipment Center.

Caro, P.W. and Isley, R. (1966). Helicopter trainee performance following synthetic flight training. Journal of the American Helicopter Society, 11(3), 38-44.

Caro, P., Isley, R. and Jolley, O. (1973). Research on synthetic training: Device evaluation and training program development (studies 1 and 2) (HumRRO Tech. Rep. 73-20). Alexandria, VA.: Human Resources Research Organization.

Caro, P., Isley, R., and Jolley, O. (1975). Mission suitability of an aircraft simulator (HumRRO Tech. Rep. 75-12). Arlington, VA: Human Resources Research Organization.

Caro, P., Shelnutt, J., and Spears, W. (1981). Aircrew training devices: Utilization (AFHRL Tech. Rep. TP-80-35). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Cohen, E. (1970). Is motion needed in flight simulators used for flight training? Human Factors, 12(1), 75-79.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin, 70, 213-220.

Condon, C., Ames, L., Hennessey, J., Shriver, E., and Seeman, R. (1979). Flight simulator maintenance training: Potential use of state-of-the-art simulation techniques (AFHRL Tech. Rep. TR-79-19). Brooks Air Force Base, TX: Air Force Human Resource Laboratory.

Cook, T.D. and Campbell, D.T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Chicago: Rand McNally College Publishing Co.

Cooper, H. (1984). The Integrative Research Review: A Social Science Approach. Beverly Hills, CA: Sage.

Cooper, H., and Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. Psychological Bulletin, 87, 442-449.

Cox, J., Wood, T., Boren, L., and Thorne, H. (1965). Functional and appearance fidelity of training devices for fixed procedures tasks (HumRRO Tech. Rep. 65-4). Alexandria, VA: Human Resources Research Organization.

Crawford, A. M., Hurlock, R. E., Padilla, R. and Sassano, A. (1976). Low cost part-task training using interavtive computer graphics for simulation of operational equipment (NPRDC-TR-TQ-76-46). San Diego, CA: Naval Personnel Research and Development Center.

Cream, W., Eggmeier, F., and Klein, G. (1978). A strategy for the development of training devices. Human Factors, 20, 145-158.

Crosby, J. (1977). Cognitive pretraining: An aid in the transition from instrument to composite flying (AFHRL Tech. Rep. TR-77-62). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Crosby, J., Pohlman, L., Leshowitz, B. and Waag, W. (1978). Evaluation of a low fidelity simulator (LFS) for instrument training (AFHRL Tech. Rep. TR-78-22). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Cyrus, M.L. (1978). Motion systems role in flight simulators for flying training (AFHRL Tech. Rep. TR-78-39, AD A059 744). Williams AFB, AZ: Flying Training Division, Air Force Human Resources Laboratory.

DeBerg, A.H., McFarland, B.P., and Showalter, T.W. (1976). The effects of simulator fidelity on engine failure training in the DC-135 aircraft. Wright-Patterson AFB, OH: Aeronautical Systems Division, pp. 83-87.

Demaree, F., Norman, D. and Matheney, W. (1965). An experimental program for relating transfer of training to pilot performance and degree of simulation (NAVTRADEVCEN Tech. Rep. 1388-1). Orlando, FL: Naval Training Device Center.

Dick, W., and Carey, L. (1985). The Systematic Design of Instruction. Glenview, IL: Scott, Foresman, and Company.

Dickinson, T.L., Hassett, C.E., and Tannenbaum, S.I. (1986). Work performance ratings: A meta-analysis of multitrait-multimethod studies (AFHRL Tech. Paper TP-86-32). Brooks AF Base, San Antonio, TX: Air Force Human Resources Laboratory, Training Systems Division.

Dederico, P. (1984). Computer-managed instruction: individual differences in student performance (NPRDC Tech. Rep. TR 84-30). San Diego, CA: Navy Personnel Research and Development Center.

Eddowes, E., and Waag, W. (1980). The use of simulator for training in-flight and emergency procedures. Neuilly Sur Seine, France: NATO Advisory Group for Aerospace Research and Development.

Edwards, B. (1986). Low-cost avionics simulation for aircrew training. (AFHRL Tech. Rep. TR-85-38). Williams Air Force Base, AZ: Air Force Human Resources Laboratory.

Edwards, B.J., Weyer, D.C., and Smith, B.A. (1979). Undergraduate pilot training: Visual discrimination pretraining for landing task (AFHRL Tech. Rep. TR-78-78). Air Force Systems Command, Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Ellis, N., Lowes, A., Matheney, W., and Norman, D. (1968). Pilot performance, transfer of training and degree of simulation: III. Performance of non-jet experienced pilots versus simulation fidelity (NAVTRADEVCEN Tech. Rep. 67-C-0334-1). Orlando, FL: Naval Training Device Center.

Eubanks, J.L. and Killeen, P.R. (1983). An application of signal detection theory to air combat training. Human Factors, 25(4), 449-456.

Federico, P. (1982). Individual differences in cognitive characteristics and computer-managed mastery learning. Personality and Individual Differences, 5, 711-724.

Feldt, L.S. (1961). The use of extreme groups to test for the presence of a relationship. Psychometrika, 26, 307-316.

Flammer, A. (1976). Does computer-assisted instruction reduce individual differences? (Tech. Rep. No. 6). Stanford University, CA: Institute for Mathematical Studies and Social Science.

Fleishman, E.A., and Quaintance, M.K. (1984). Taxonomies of Human Performance. New York: Academic Press.

Flexman, R., Roscoe, S., Williams, A., and Williges, B. (1972). Studies in pilot training: The anatomy of transfer. Aviation Research Monographs, 2. Urbana-Champaign, Il: Aviation Research Laboratory.

Fuller, J., Waag, W., and Martin, E. (1980). Advanced simulator for pilot training: Design of automated performance measurement system (AFHRL Tech. Rep. TR-79-57). Williams Air Force Base, AZ: Air Force Human Resource Laboratory.

Gagne', R.M. (1985). Conditions of Learning (fourth Ed.). New York: Holt, Rinehart, & Winston.

Gerathewohl, S. (1969). Fidelity of simulation and transfer of training: A review of the problem (Rep. No. FAA-AM-69-24). Washington, D.C.: Federal Aviation Administration.

Glass, G.V., McGaw, B., and Smith, M.L. (1981). Meta-analysis in Social Research. Beverly Hills, CA: Sage.

Glass, G.V. and Stanley, J.C. (1970). Statistical Methods in Education and Psychology. Englewood Cliffs, N.J. Prentice-Hall.

Goebel, T., Baum, D., and Hagin, W. (1971). Using a ground trainer in a job sample approach to predicting pilot performance (AFHRL Tech. Rep. TR-80-3). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Goodman, H.J. (1978). "Cognitive mapping", "learning styles", and "sensory modality preferences" as factors in individualized instruction: A position paper on the as yet largely untapped research potential of integrated information systems when combined with educational technology. Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, Canada.

Gray, T. (1979). Boom operator part task trainer: Test and evaluation of the transfer of training (AFHRL Tech. Rep. TR-79-37). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Gray, T.H., Chun, E.K., Warner, H.D., and Eubanks, J.L. (1981). Advanced flight simulator: Utilization in A-10 conversion and air-to-surface attack training (AFHRL Tech. Rep. TR-80-20). Air Force Systems Command, Brooks AFB, TX: Air Force Human Resources Laboratory.

Gray, T.H. and Fuller, R. R. (1977). Effects of simulator training and platform motion in air-to-surface weapon delivery training (AFHRL Tech. Rep. TR-77-29, AD043 649). Williams AFB, AZ: Flying Training Division, Air Force Human Resources Laboratory.

Grunwald, W. (1968). An investigation of the effectiveness of training devices with varying degrees of fidelity. Unpublished doctoral dissertation, University of Oklahoma. (University Microfilms No. 68-13,559).

Gum, D., Albery, W., and Basinger J. (1975). Advanced simulation in undergraduate pilot training: An overview (AFHRL Tech. Rep. TR-75-59(1)). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Gundry, J. (1976). Man and motion cues. Third Flight Simulation Symposium Proceedings, Royal Aeronautical Society, London.

Hagin, W., Durall, E., and Prophet, W. (1979). Transfer of training evaluation: U.S. Navy device 2B35 (Seville Tech. Rep. 79-06). Pensacola, FL: Seville Research Corporation.

Hagin, W.V. and Smith, J.F. (1974). Advanced simulation in undergraduate pilot training (ASUPT) facility utilization plan (AFHRL Tech. Rep. TR-74-43, AD786 411). Williams AFB, AZ: Flying Training Division, Air Force Human Resources Laboratory.

Hammerton, M. (1967). Measures for the efficiency of simulators as training devices. Ergonomics, 10(1) 63-65.

Hays, R. (1980). Simulator fidelity: A concept paper (ARI Tech. Rep. 490). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Hays, R. (1981). Training simulator design guidance: The iterative data base approach (ARI Tech. Rep. 545). Alexandria, VA: U.S. Army Research Institute.

Hays, R. (1981). The ARI fidelity workshop: Summary comments. In R.T. Hays (Ed). Research issues in the determination of simulator fidelity: Proceedings of the ARI sponsored workshop, July 23-24, 1981 (ARI Tech. Rep. 547). Alexandria, VA: U.S. Army Research Institute.

Hays, R.T., and Carson, R. (1988). Development of an automated training specification system (ATSS). In A.L. Smode (Ed.), Independent Research and Independent Exploratory Development (ID/IED) Programs Annual Report FY87, Orlando, FL: Naval Training Systems Center.

Hays, R.T. and Singer, M. J. (1983). Research issues in training device design: The organization of a data base (ARI Tech. Rep. 588) Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Hays, R.T. and Singer, M.J. (1989). Simulation Fidelity in Training System Design: Bridging the Gap Between Reality and Training. New York: Springer-Verlag.

Hays, W.L. (1973). Statistics for the Social Sciences. New York: Holt, Rindhart and Winston.

Holman, G.L. (1979). Training effectiveness of the CH-47 flight simulator (ARI Res. Rep. 1209). Alexandria, VA: U.S. Army Institute.

Hogan, J., Arneson, S., and Salas, E. (1987). Individual differences in military training environments: Four areas of research (NAVTRASYSCEN Tech. Rep. TR87-003). Orlando, FL: Human Factors Division, Naval Training Systems Center.

Holding, D.H. (1962). Transfer between difficult and easy tasks. British Journal of Psychology, 53, 397-407.

Hopkins, C. (1975). How much should you pay for that box? Human Factors, 17, 533-541.

Huff, E.M. and Nagel, D.C. (1975, March). Psychological aspects of aeronautical flight simulation. American Psychologist, pp. 426-438.

Hughes, R. (1979). Advanced training features: Bridging the gap between in-flight and simulator based models of flying training (AFHRL Tech. Rep. TR-78-96). Williams Air Force Base, AZ: Air Force Human Resource Laboratory.

Hughes, R., Hannan, S. and Jones, W. (1979). Application of flight simulator record/playback feature (AFHRL Interim Rep. 79-52). Williams Air Force Base, AZ: Air Force Human Resources Laboratory.

Hunter, J.E., Schmidt, F.L., and Jackson, G.B. (1982). Meta-analysis. Cumulating Research Findings Across Studies. Beverly Hills, CA: Sage.

Hutton, D.P., Burke, D.K., Englehart, J.D., Wilson, J.M., Romaglia, F.J. and Schneider, A.J. (1976). Air to ground visual simulation demonstration. (Final Report, Project 2235). Wright Patterson AFB, OH: Simulator S.P.O.'s, Aeronautics System Division.

Ince, F., Williges, R.C. and Roscoe, S.N. (1975). Aircraft simulator motion and the order merit of flight attitude and steering guidance displays. Human Factors, 17(4), 388-400.

Irish, P. and Buckland, G. (1978). Effects of platform motion, visual and g-seat factors upon experienced pilot performance in the flight simulator (AFHRL Tech. Rep. TR-78-9). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Isaac, S., and Michael, W.B. (1978). Handbook in Research Evaluation (10th Ed.). San Diego, CA: Edits Publishers.

Isley, R., Caro, P. and Jolley, O. (1968). Evaluation of synthetic instrument flight training in the officer/warrant officer rotary wing aviatory course (HumRRO Tech. Rep. 68-14). Alexandria, VA: Human Resources Research Organization.

Jackson, G.B. (1980). Methods for integrative reviews. Review of Educational Research, 50, 438-460.

Jacobs, R.S. (1975). Simulator motion as a factor in flight simulator effectiveness. U.S. Department of Health, Education and Welfare Education Resources Information Center Report.

Jacobs, R.S. (1976). Simulator cockpit motion and the transfer of initial flight training (ARL Tech. Rep. 76-8/AFOBA Tech. Rep. 76-4). Aviation Research Laboratory, Institute of Aviation, University of Illinois at Urbana-Champaign.

Jacobs, R.S. and Roscoe, S.N. (1975). Simulator cockpit motion and the transfer of initial flight training (ARL Tech. Rep. 75-18/AFOSR Tech. Rep. 75-8). Aviation Research Laboratory, Institute of Aviation, University of Illinois at Urbana-Champaign.

Jacobs, R., Williges, R. and Roscoe, S. (1973). Simulator motion as a factor in flight-director display evaluation. Human Factors, 15, 569-582.

Johnson, S.L. and Roscoe, S.N. (1972). What moves, the airplane or the world? Human Factors, 14(2), 107-129.

Johnson, S., Williges, R. and Roscoe, S. (1971). A new approach to motion relations for flight director displays (ARL Tech. Rep. 71-20). Savoy, IL: University of Illinois, Institute of Aviation Research Laboratory.

Jones, A.P., Johnson, L.A., Butler, M.C., and Main, D.S. (1983). Apples and oranges: An empirical comparison of commonly used indices of interrater agreement. _Academy of Management Journal_, _26_, 507-519.

Jones, E.R., Hennessy, R.T. and Deutsch, (Eds). (1985). _Human factors aspects of simulation_. Working Group on Simulation, National Research Council, National Academy Press, Washington, D.C.

Jones, M.B., Kennedy, R.S., Turnage, J.J., Kuntz, L.A., and Jones, S.A. (1986). _Meta-analysis of human factors engineering studies comparing individual differences, practice effects, and equipment design variations_ (Final report for contract #F33615-85-C-0539). Wright-Patterson AFB, OH: Aeronautical Systems Division/PMRSB, Air Force Systems Command.

Kendall, M.G. and Stuart, A. (1961-1966). _The Advanced Theory of Statistics (Volumes 1-3)_. London: Griffin.

Kennedy, R.S., Bittner, A.C., and Jones, M.B. (1981). Video-game and conventional tracking. _Perceptual Motor Skills_, _53_, p. 310.

Kennedy, R.S., Wulfeck, J.W., Prosin, D.J. and Burger, W.J. (1974). _Effect of a predictor display on carrier landing performance_ (TP-74-46). Naval Miss le Center, Point Mugu, CA: Human Factors Engineering Branch, System Integration Division.

Kinkade, R. and Wheaton, G. (1972). Training device design. In H. Van off and R. Kinkade (Eds). _Human engineering guide to equipment design_. Washington, D.C.: American Institutes for Research.

Klein, G.A., and Brezovic, C.P. (1987). _Human Performance Data Needed for Training Device Design Decisions_ (AAMRL Tech. Rep. TR-87-010). Wright-Patterson AFB, OH: Harry G. Armstrong Aerospace Medical Research Laboratory, Human Systems Division.

Koonce, J. (1974). _Effects of ground-based aircraft simulator conditions upon predictions of pilot proficiency_ (ARL Tech. Rep. 74-5). University of Illinois at Urbana-Champaign: Aviation Research Laboratory.

Koonce, J.M. (1979). Predictive validity of flight simulators as a function of simulator motion. _Human Factors_, _21_(2), 215-223.

Kotora, J. and Giebert, W. (1982). _Training effectiveness evaluation of device AF37A-T59, C-130E instrument flight simulator (IFS)_. Proceedings of the 4th Interservice/Industry Training Equipment Conference, Orlando, FL.

Krahenbuhl, G., Marett, J. and Reid, G. (1978). simulator pretraining and in flight stress of student pilots. _Aviation, Space and Environmental Medicine_, 1007-1110.

Kron, G. (1970). Fidelity of aircraft simulation (Link IR&D Tech. Rep. LR-439). Binghamton, NY: Singer Company.

Kruskal, W.H. (1958). Ordinal measures of association. Journal of the American Statistical Association, 53, 814-861.

Kulhavey, R.W. (1977). Feedback in written instruction. Review of Educational Research, 47, 211-232.

Kulhavey, R.W., White, M.T., Topp, B.W., Chan, A.L., and Adams, J. (1985). Feedback complexity and corrective efficiency. Contemporary Educational Psychology, 10, 285-291.

Lintern, G. (1980). Transfer of landing skill after training with supplementary visual cues. Human Factors, 22, 215-223.

Lintern, G., Kaul, C.E. and Collyer, S.C. (1984). Glideslope descent-rate cuing to aid carrier landings. Human Factors, 26(6), 667-675.

Lintern, G., and Kennedy, R.S. (1982). A video game as a covariate for carrier landing research. Manuscript submitted to the Eighth Psychology in the DoD Symposium.

Lintern, G., Thomley, K., Nelson, B. and Roscoe, S. (1984). Content, variety and augmentation of simulated visual scenes for teaching air-to-ground attack (NAVTRAEQUIPCEN Tech. Rep. 81-C-0105-3). Orlando, FL: Naval Training Equipment Center.

Martin, E. (1981). Training effectiveness of platform motion: Review of motion research involving the advanced simulator for pilot training and the simulator for air-to-air combat (AFHRL Tech. Rep. TR-79-56). Williams Air Force Base, AZ: Air Force Human Resources Laboratory.

Martin, E. and Cantaneo, D. (1980). Computer generated image: Relative training effectiveness of day versus night visual scenes (AFHRL Tech. Rep. TR-79-56). Williams Air Force Base, AZ: Air Force Human Resources Laboratory.

Martin, E.L. and Rinalducci, E.J. (1983). Low-level flight simulation: Vertical cues (AFHRL Tech. Rep. TR-83-17). Williams AFB, AZ: Air Force Human Resources Laboratory Operations Training Division.

Martin, E. and Waag, W. (1978a). Contributions of platform motion to simulator training effectiveness: Study I--basic contact (AFHRL Tech. Rep. TR-78-15). Williams Air Force Base, AZ: Air Force Human Resources Laboratory.

Martin, E.L. and Waag, W.L. (1978b). Contributions of platform motion to simulator training effectiveness: Study II Basic Contact. (AFHRL Tech. Rep. TR-78-52, AD A064 305). Williams Air Force Base, AZ: Flying Training Division, Air Force Human Resources Laboratory.

Mathieu, J. and Tannenbaum, S.I. (1983, March). Analyzing the analysis today. In S.I. Tannenbaum & E. Salas (Co-chairs), Meta-analysis in I/O and O.B.: Probes of the Technique, Current Practices, and the Future. Symposium presented at the I/O and O.B. Graduate Student Convention, Chicago, IL.

McDaniel, W.C., Scott, P.G. and Browning, R. F. (1983). Contribution of platform motion simulation in SH-3 helicopter pilot training (TAEG Rep. 153). Orlando, FL: Training Analysis and Evaluation Group.

McGaugh, M.F. and Holman, G.L. (1977). Operational test II of CH47C Synthetic Flight Training System (2B31)(SFTS) (AD-B022 470). Fort Rucker, AL: United States Army Aviation Board.

McGuiness, J., Bowman, J.H. and Puig, J.A. (1982). Effectiveness evaluation for air combat training, Person-System Integration. Proceedings of the 4th Interservice/Industry Training Equipment Conference, Orlando, FL.

Melden, T.M. and Houston, R.C. (1975). Major changes in flight crew training: 8 years of experience at American Airlines (AIAA Paper No. 75-1049) Aircraft Systems and Technology Meeting, Los Angeles, CA., August 4-7, 1975. American Institute of Aeronautics and Astronautics, 1290 Avenue of the Americas, New York, N.Y. 10019.

Micheli, G. (1972). Analysis of the transfer of training, substitution and fidelity of simulation of training equipment (TAEG Tech. Rep. 2). Orlando, FL: Naval Training Equipment Center.

Miller, G. (1974). Some considerations in the design and utilization of simulators for technical training (AFHRL Tech. Rep. TR-74-65). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Miller, R. (1953). Handbook on training and training equipment design (WADC Tech. Rep. 53-136). Wright Patterson Air Force Base, OH: Wright Air Development Center.

Montemerlo, M.D. and Tennyson, M.E. (1976). Instructional Systems Development: Conceptual Analysis and Comprehensive Bibliography (NAVTRAEQUIPCEN Tech. Rep. IH-257). Orlando, FL: Naval Training Equipment Center.

Mudd, S. (1968). Assessment of the fidelity of dynamic flight simulators. Human Factors, 10, 351-358.

Mullen, B., and Rosenthal, R.   (Eds). (1985).   Basic meta-analysis: Procedures and Programs. Hillsdale, NJ: Lawrence Erlbaum.

Nataupsky, M., McCloy, T., Bermudez, J. and Tirman, V.   (1980).   Effects of criteria on flight simulation: Study II - multiple criteria (Frank J. Seiler Research Laboratory Final Report 80-0020). Boulder, CO: U.S. Air Force Academy.

Nataupsky, M., Waag, W., Weyer, D., McFadden, R. and McDowell, E.   (1979). Platform motion contributions to simulator training: Study III--interaction of motion with field-of-view (AFHRL Tech. Rep. TR-79-25). Williams Air Force Base, AZ: Air Force Human Resources Laboratory.

Nauta, F. (1985).   Alleviating Fleet Maintenance Problems through Maintenance Training and Aiding Research (NAVTRAEQUIPCEN Tech. Rep. MDA903-81-C-0166-1). Orlando, FL: Naval Training Equipment Center.

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

Orlansky, J. and String, J.   (1977).   Cost-effectiveness of flight simulators for military training. Volume I: Use and effectiveness of flight simulators (IDA Paper P-1275). Arlington, VA: Institute for Defense Analyses.

Orwin, T. and Cordray, D.   (1985).   Effects of deficient reporting on meta-analysis: A conceptual framework and reanalysis. Psychological Bulletin, 97, 134-137.

Osborne, S., Broyles, B. and Quick, K. (1985).   Training effectiveness evaluation of device 2F110 E-2C aircraft operational flight trainer (NAVTRASYSCEN Tech. Rep. TR-85-002). Orlando, FL: Naval Training Systems Center.

Parris, B. and Cook, A.   (1978).   Effects of visual and motion cueing systems on pilot performance during takeoffs with engine failures (NASA Tech. Paper 1365). Moffett Field, CA: National Aeronautics and Space Administration.

Payne, T.A., Hirsch, D.L., Semple, C.A., Farmer, J.R., Spring, W.G., Sanders, M.S., Wimer, C.A., Carter, V.E. and Hu, A.   (1976).   Experiments to evaluate advanced flight simulation in air combat pilot training. (NOR-76-52). Warmister, PA: Naval Air Development Center.

Payne, D.L., and Tirre, W.C. (1983).   Individual differences in learning rate. Paper presented at the 9th Annual Department of Defense Symposium, Colorado Springs, CO.

71

Pearlman, K., Schmidt, F.L., and Hunter, J.E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology*, *65*, 373-406.

Pfeiffer, M.G., & Horey, J.D. (1987). *Training effectiveness of aviation motion simulation: A review and analysis of the literature* (Spec. Rep. 87-007). Orlando, FL: Naval Training Systems Center.

Pierce, B.J. (1983). *Low altitude simulator training: A-10 aircraft* (AFHRL Tech. Paper TP-82-43). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Pohlman, L. and Reed, J. (1978). *Air to air combat skills: Contributions of platform motion to initial training* (AFHRL Tech. Rep. TR-78-53). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Povenmire, H.K. and Roscoe, S. (1971). An evaluation of ground-based flight trainers in routine primary flight training. *Human Factors*, *13*, 109-116.

Prather, D.C., Berry, G.A. and Jones, G.L. (1971). *Four in-depth studies of a flight training skill: Trial and error versus prompted learning evaluated on efficiency, transfer and stress* (Rep. 71-11). Fort Collins, CO: U.S. Air Force Academy.

Prophet, L. and Boyd, A. (1970). *Device task fidelity and transfer of training: Aircraft cockpit procedures training* (AFHRL Tech. Rep. TR-78-53). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Puig, J., Harris, W. and Ricard, G. (1978). *Motion in flight simulation: An annotated bibliography* (NAVTRAEQUIPCEN Tech. Rep. IH-298). Orlando, FL: Naval Training Equipment Center.

Randle, R., Tanner, T., Hamerman, J. and Showalter, T. (1981). *The use of total simulator training in transitioning air-carrier pilots: A field evaluation* (NASA Tech. Memo. 81250). Moffett Field, CA: National Aeronautics and Space Administration.

Reed, J. and Reed, J. (1979). *Air refueling director lights trainer: Analysis and evaluation of training effectiveness* (AFHRL Interim Rep. 78-75). Williams Air Force Base, AZ: Air Force Human Resource Laboratory.

Reicher, G., Davidson, B., Hawkins, H. and Osgood, G. (1980). *Low cost simulation of piloting tasks* (ONR Tech. Rep. 4). Arlington, VA: Office of Naval Research.

Reid, G. and Cyrus, M. (1974). _Transfer of training with formation flight trainer_ (AFHRL Tech. Rep. TR-74-102), Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Reid, G.B. and Cyrus, M.L. (1977). _Formation flight trainer evaluation of the T-37 UPT_ (AFHRL Tech. Rep. TR-77-23). Williams Air Force Base, AZ: Flying Training Division, AFHRL, (AD-A043 197).

Ricard, G. and Parrish, R. (1984). Pilot differences and motion cuing effects on simulated helicopter hover. _Human Factors_, _26_, 249-256.

Ricard, G.L., Parrish, R.V., Ashworth, B.R. and Wells, M.D. (1981). _The effects of various fidelity factors on simulated helicopter hover_ (Tech. Rep. NAVTRAEQUIPCEN-IH-321). Orlando, FL: Naval Training Equipment Center.

Rolfe, J. and Waag, W. (1980). _Fidelity of flight simulation for pilot training_. (Royal Air Force Education Bulletin 18).

Roscoe, S.N. (1980). _Aviation Psychology_. The Iowa State University Press/Ames.

Roscoe, S., Hasler, S. and Dougherty, D. (1966). Flight by periscope: Making takeoffs and landings; the influence of image magnification, practice, and various conditions of flight. _Human Factors_, _17_, 13-40.

Roscoe, S. and Williges, R. (1975). Motion relationships in aircraft attitude and guidance: A flight experiment. _Human Factors_, _17_, 374-387.

Rose, A.M., Wheaton, G.R. and Yates, L.G. (1985). _Forecasting device effectiveness issues: Volume I_ (ARI Tech. Rep. 680). Alexandria, VA: Army Research Institute.

Rosenthal, R. (1978). Combining results of independent studies. _Psychological Bulletin_, _85_, 185-193.

Rosenthal, R. (Ed.). (1980). _New Directions for Methodology of Social and Behavioral Science: Quantitative Assessment of Research Domains_ (No. 5). San Francisco, CA: Jossey-Bass.

Rosenthal, R. and Rosnow, R. (1984). _Essentials of behavioral research: Methods and data analysis_. McGraw-Hill, Inc.

Rouse, W.B. (1983). On better mousetrap and basic research: Getting the applied world to the laboratory door. _IEEE Systems, Man, and Cybernetics Conference_.

Royer, J.M. (1979). Theories of transfer. _Educational Psychologist_, _14_, 53-69.

Ruocco, J., Vitale, P. and Benfari, R. (1965). Kinetic cueing in simulated carrier approaches (NAVTRADEVCEN Tech. Rep. 1432-1). Port Washington, NY: Naval Training Device Center.

Ryan, L.E., Puig, J.A., Michelli, G.S. and Clarke, J.C. (1972). An evaluation of the training effectiveness of device 2F90, TA-4J operational flight trainer, Part K: B stage (NAVTRAEQUIPCEN IH 207). Orlando, FL. August 1972.

Ryan, L., Scott, P. and Browning, R. (1978). The effects of simulator landing practice and the contribution of motion simulation to P-3 pilot training (TAEG Tech. Rep. 63). Orlando, FL: Naval Training Device Center.

Schmitt, N., Gooding, R.Z. Noe, R.A., and Kirsch, M. (1984). Meta-analysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. Personnel Psychology, 37, 407-422.

Seevers, J. and Makinney, R. (1979). Simulator for air-to-air combat motion systems investigation (AFHRL Tech. Rep. 79-18). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Semple, C. (1981). Fidelity: What we know and who should know it. In R. Hays (Ed), Research issues in the determination of simulator fidelity: Proceedings of the ARI sponsored workshop, July 23-24, 1981 (ARI Tech. Rep. 547). Alexandria, VA: U.S. Army Research Institute.

Semple, C., Hennessy, R., Sanders, M., Cross, B., Beith, B. and McCauley, M. (1981). Aircrew training devices: Fidelity features (AFHRL Tech. Rep. TR-80-36). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Showalter, T. and Parris, B. (1980). The effects of motion and g-seat cues on pilot simulator performance of three piloting tasks (NASA Tech. Paper 1601). Moffett Field, CA: Ames Research Center.

Shrout, P.E., and Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86, 420-428.

Simon, C. and Roscoe, S. (1984). Application of a multifactor approach to transfer of training research. Human Factors, 26, 591-612.

Skinner, B.F. (1953). Science and Human Behavior. New York: McMillan.

Smith, R.L., Pence, G.G., Queen, J.E., and Wulfeck, J.W. (1974). Effect of a predictor instrument on learning to land a simulated jet trainer. Dunlap and Associates, Inc. Western Division, 115 South Oak Street, Inglewood, CA.

Smith, B., Waters, B. and Edwards, B. (1975). Cognitive pretraining of the TO37 overhead traffic pattern (AFHRL Tech. Rep. 75-72). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Smode, A. (1971). Human factors inputs to the training device design process (NAVTRAEQUIPCEN Tech. Rep. 69-C-02980-1). Orlando, FL: Naval Training Equipment Center.

Smode, A. (1972). Training device design: Human factors requirements in the technical approach (NAVTRAEQUIPCEN Tech. Rep. 71-C-0012-1). Orlando, FL: Naval Training Device Center.

Smode, A. and Hall, E. (1975). Translating information requirements into training device fidelity requirements. Proceedings of the 19th Annual Meeting of the Human Factors Society. 33-36.

Smode, A., Hall, E. and Meyer, D. (1966). An assessment of research relevant to pilot training (AMRL Tech. Rep. 66-196). Wright Patterson Air Force Base, OH: Aerospace Medical Research Laboratories.

Spears, W., Sheppard, H., Roush, M. and Richetti, C. (1981). Simulator training requirements and effectiveness study: Abstract bibliography (AFHRL Tech. Rep. TR-80-38). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Strube, M.J. and Hartman, D.P. (1983). Meta-analysis: Techniques, applications, and functions. Journal of Consulting and Clinical Psychology, 51, 14-27.

Su, D. (1984). A review of the literature on training simulators: Transfer of training and simulator fidelity (Res. Rep. #84-1). Washington, DC: Office of Naval Research.

Thorndike, E.L. (1903). Educational Psychology. New York: Lemcke and Buechner.

Thorpe, J., Varney, N., McFadden, R., LeMaster, W. and Short, L. (1978). Training effectiveness of three types of visual systems for KC-135 flight simulators (AFHRL Tech. Rep. TR-78-16). Williams Air Force Base, AZ: Air Force Human Resources Laboratory.

Tinsely, H.E.A., and Weiss, D.J. (1975). Interrater reliability and agreement of subjective judgements. Journal of Counseling Psychology, 22, 358-376.

Tracey, W.R. (1984). Designing Training and Development Systems. New York: American Management Association (AMACOM).

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

Waag, W. (1981). _Training effectiveness of visual and motion simulation_ (AFHRL Tech. Rep. TR-79-72). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Weitzman, D., Fineberg, M., Gade, P. and Compton, G. (1979). Proficiency maintenance and assessment in an instrument flight simulator. _Human Factors_, _21_, 701-710.

Weitz, J. and Adler, S. (1973). The optimal use of simulation. _Journal of Applied Psychology_, _58_(2), 291-224.

Weller, D.R. (1983). _Predictor displays as training aids in carrier landings_. (Tech. Note NAVTRAEQUIPCEN TN-66). Human Factors Laboratory, Orlando, FL: Naval Training Equipment Center.

Westra, D.P. (1982). _Simulator design features for carrier landings: II. Indsimulator transfer of training_ (NAVTRAEQUIPCEN Tech. Rep. 81-C-0105-1). Orlando, FL: Naval Training Equipment Center.

Westra, D.P. and Lintern, G. (1985). _Simulator design features for helicopter landing on small ships. I. A performance study_ (NAVTRASYSCEN Tech. Rep. 81-C-0105-13). Orlando, FL: Naval Training Systems Center.

Westra, D.P., Lintern, G., and Wightman, D.C. (1986). _Aircraft carrier landing research at the Visual Technology Research Simulator (VTRS): Implications for simulator design_. Paper presented at the Eastern Simulation Conference/Society for Computer Simulation, March 10-12.

Westra, D.P., Lintern, G., Sheppard, D.J., Thomley, K.E., Mauk, R., Wightman, D.C. and Chambers, W.S. (1986). _Simulator design and instructional features for carrier landing: Transfer study_ (NAVTRASYSCEN Tech. Rep. TR-85-C-0044-2). Orlando, FL: Naval Training Systems Center.

Wheaton, G., Rose, A., Fingerman, P., Korotkin, A., Holding, D. and Mirabella, A. (1976). _Evaluation of the effectiveness of training devices: Literature review and preliminary model_ (ARI Res. Memo. 76-6). Alexandria, VA: U.S. Army Research Institute.

Wightman, D.C., and Lintern, G. (1985). Part-task training for tracking and manual control. _Human Factors_, _27_(3), 267-283.

Wightman, D.C., and Sistrunk, F. (1987). Part-task training strategies in simulated carrier landing final-approach training. _Human Factors_, _29_(3), 245-254.

Williges, B., Roscoe, S. and Willeges, R. (1973). Synthetic flight training revisited. _Human Factors_, _15_, 543-560.

Willson, V.L. (1976). Critical values of the rank biserial correlation coefficient. _Educational and Psychological Measurement_, _36_, 297-300.

Winer, B.J. (1971). _Statistical Principles in Experimental Design_ (Second Edition). New York: McGraw-Hill.

Wolf, F.M. (1986). _Quantitative methods for research synthesis_. Beverly Hills, CA: Sage Publications.

Wood, M.E., Hagin, W.V., O'Connor, R. and Myers, D.W. (1972). _Design of the simplified formation trainer_ (AFHRL Tech. Rep. TR-72-8). Williams Air Force Base, AZ: Flying Training Division, AFHRL. (AD-A754 973)

Woodruff, R.R. (1979). _Effects of varying visual display characteristics of the T-4G, a T-37 flight simulator_ (AFHRL Tech. Rep. TR-79-19). Williams Air Force Base, AZ: Air Force Human Resources Laboratory, Flying Training Division.

Woodruff, R., Hubbard, D. and Shaw, A. (1985). _Advanced simulator for pilot training and helmet-mounted visual display configuration comparisons_. (AFHRL Tech. Rep. TR-84-65). Williams Air Force Base, AZ: Air Force Human Resources Laboratory, Operations Training Division.

Woodruff, R.R. and Smith, J.F. (1974). _T-4G simulator and T-4 ground training devices in U.S.A.F. undergraduate pilot training: An exploratory study_ (AFHRL Tech. Rep. TR-74-78). Williams Air Force Base, AZ: Flying Training Division, AFHRL. (AD-A008 197).

Woodruff, R., Smith, J., Fuller, J. and Weyer, D. (1976). _Full mission simulation in undergraduate pilot training: An exploratory study_ (AFHRL Tech. Rep. TR-76-84). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Young, L., Jensen, R. and Treichel, C. (1973). _Uses of a visual landing system in primary flight training_. Urbana-Champaign: Univ of Illinois, Institute of Aviation. (AD-784 888).

Yuan-Liang, D. (1984). _A review of the literature on training simulators: Transfer of training and simulator fidelity_ (GATECH Tech. Rep. 84-1). Atlanta, GA: Georgia Institute of Technology.

APPENDIX A

Experiments Excluded from the Meta-analysis

AUTHOR (YR)	REASON FOR EXCLUSION
Bailey, Hughes & Jones (1980)	No Transfer
Biersner (1976)	No Transfer
Billings, Gerke & Wick (1975)	No Training
Brictson & Briendenback (1981)	Insufficient Statistics
Browning, McDaniel & Scott (1982)	Insufficient Statistics
Burger & Brictson (1976)	No Training or Transfer
Caro, Isley & Jolley (1973)	Insufficient Statistics
Caro, Isley & Jolley (1975)	Insufficient Statistics
Crawford, Hurlock, Padilla & Sassano (1976)	No Transfer
Crosby (1977)	No Transfer
Demaree, Norman & Matheney (1965)	No Training or Transfer
Edwards, Weyer & Smith (1979)	No Transfer
Ellis, Lowes, Matheney & Norman (1968)	No Training or Transfer
Hagin, Duvall & Smith (1979)	Insuficient Statistics
Ince, Williges & Roscoe (1975)	No Transfer
Irish & Buckland (1978)	No Training or Transfer
Jacobs & Roscoe (1975)	Not Appropriate Statistics
Jacobs, Williges & Roscoe (1973)	No Transfer
Koonce (1979)	Insufficient Statistics
Krahenbuhl, Marett & Reid (1978)	Inappropriate Measures
Lintern (1980)	Fixed Wing Not Jet
Prather, Berry & Jones (1971)	No Transfer
Povenmire & Roscoe (1971)	Insufficient Statistics
Prophet & Boyd (1970)	Insufficient Statistics
Reicher, Davidson, Hawkins & Osgood (1980)	Insufficient Statistics
Reid & Cyrus (1974)	Insufficient Statistics
Reid & Cyrus (1977)	Insufficient Statistics
Roscoe & Williges (1975)	No Training
Ruocco, Vitale & Benfari (1965)	No Transfer
Ryan, Scott & Browning (1978)	Insufficient Statistics
Thorpe, Varnesey, McFadden, Lemaster & Short (1978)	Insufficient Statistics
Smith, Pence, Queen & Wulfek (1974)	No Transfer
Woodruff & Smith (1974)	Insufficient Statistics
Woodruff, Smith, Fuller & Weyer (1976)	Insufficient Statistics
Woodruff, Smith & Harris (1974)	Insufficient Statistics
Young, Jensen & Treichel (1973)	Insufficient Statistics

APPENDIX B
PRELIMINARY CODE SHEET WITH DESCRIPTION OF AREA TOPICS

HELICOPTER

JET

(circle one)

EXPERIMENTAL COMPARISON:

_____ 1. Simulator Training vs. Aircraft-only Training
_____ 2. Motion vs. No Motion
_____ 3. Limited vs. Full Amount of Simulator Training
_____ 4. Other Comparison (describe)

STUDY ID #: _____

CODER ID: _____

DEVICE NAME: _____

AIRCRAFT: _____

1. REPORTED
   STATISTIC
   VALUE(S)

2. CORRESPONDING
   VALUES:
   MEANS, SD's,
   & N's

Identifying information:
dep. measure, group
comparisons, etc.

Study Characteristics	Subject Characteristics	Training Length/ Criteria	Task(s) Trained
1. Subj. assignment	1. Service branch	1. Total number hours/trials	1. Total number tasks
2. Use of matching	2. Prior experience	2. Criteria for advancement	2. Task classification (if applicable)
3. Study comparisons		3. Rating system used	3. Task names by classification (if applicable)
4. Dep. measure(s)		4. Inter-rater reliability	
5. Type of statistics reported			

COMMENTS/PROBLEMS:

1. Design features, such as counterbalancing IP-student pairings for training and assessment, and IP's blind to group assignment during evaluation phase.

2. Any irregularities that would affect internal or external validity of study results.

APPENDIX C

ITEMS DELETED FROM FINAL CODE SHEET DUE TO LACK OF
INFORMATION/VARIABILITY IN AIRCRAFT SIMULATOR
TRAINING EXPERIMENTS BY CODING AREA

Training Characteristics

What was the instructor-student ratio?

What was the instructor's reported level of acceptance for use of simulator as a training device?

What was the student's reported level of acceptance for use of simulator as a training device?

To what extent did training incorporate ISD principles and procedures?

Was knowledge of results (KOR) given to student?

If KOR was given, in what form was it given (i.e., system versus instructor generated, etc.)?

Was there any attempt to transition students from simulator to aircraft?

Were IPs trained to use instructional features of simulator?

Were any part-task training methods employed in conjunction with simulator training features?

Simulator Fidelity Characteristics

Motion system:
   *a) Was a stick shaker or buffet system used?

Sound system:
   *a) Was there any sound simulation used?

Cockpit display and flight control characteristics:
   *a) Was cockpit (instrumentation/controls) PHYSICALLY similar to the transfer aircraft?

   *b) Does instrumentation/controls FUNCTIONALLY represent that in transfer aircraft?

*c) Was joystick "feel" similar to that of aircraft?

b) Were the simulator flight characteristics validated using actual aircraft data?

Training features:
a) What special environmental features were used?

b) What special training features were used?

* Indicates item excluded due to lack of variability in experiments used in analyses; all others were excluded due to lack of information (i.e., not stated in report).

Trainee Characteristics and Study Design

Was a covariate used to reduce error variance in performance measures?

If a covariate was used, was it cognitive or non-cognitive (e.g., personality assessment, physical ability, etc.) in nature?

For training, were instructor pilot (IP)-subject pairings counterbalanced?

For assessment, were IP-subject pairings counterbalanced?

Were IP's blind to type training given students?

If IP ratings were used to assess student performance, what was the reported inter-rater reliability estimate?

If IP ratings were used, what was the reported intra-rater reliability estimate?

* How did the study assess student's prior flight experience (paperpencil test, UPT grades, background check, etc.)?

To what extent was the training program the same across treatments?

* Indicates item excluded due to lack of variability in experiments used in analyses; all others were excluded due to lack of information (i.e., not stated in report).

APPENDIX D

CODE SHEET WITH RESEARCH FREQUENCIES FOR RESPONSE
CATEGORIES AND MEANS FOR CONTINUOUS ITEMS

An asterisk (*) on an item indicates that an experiment may be coded in more than one category on that item.

RESEARCH IDENTIFICATION

1. Experiment ID #: ___N/A

2. Coder ID #: ___N/A___

3. Date of publication: ___N/A___

4. Publication Source:

    __0__ Book   __0__ Dissertation

    __0__ Journal   __0__ Paper Presentation

    __19__ Technical Report   __0__ Unpublished Manuscript

    __0__ Other (describe) _____

SIMULATOR/AIRCRAFT INFORMATION

5. Simulator name/identification code: _____N/A_____

6. Aircraft name/identification code: _____N/A_____

STUDY DESIGN AND SUBJECT CHARACTERISTICS

7. Subject assignment to groups was:

    __5__ Random Only   __10__ Matching Only

    __3__ Neither   __1__ Unstated/Unclear

8. Subjects were:

    __8__ Recent Undergraduate Pilot Training (UPT) Graduates

    __0__ Experienced Pilots Transitioning to New Aircraft

    __2__ Mixed - Having Both High And Low Experienced Pilots

___9__ Other (describe) <u>UP TRAINEES (N=6); RECENT AF ACADEMY GRADS; F-14 TRAINEES ENTERING AIR REFUELING STAGE; FCLP TRAINEES</u>

*9. Group contrasts consisted of:

___10__ Simulator Training Versus Aircraft Training (Control)
(SIMULATOR TRAINING vs FLY ONLY CONTROL)

___5___ Simulator Training With Motion System On Versus Simulator Training With Motion System Off
(MOTION vs NO MOTION)

___2___ Full Amount Of Simulator Training Versus Limited Amount Of Simulator Training (FULL vs LIMITED TRAINING)

___2___ OTHER (describe) <u>OFT + CFT VS OFT ACONE;   OLD VS.   NEW SIM.</u>

SIMULATOR FIDELITY CHARACTERISTICS

10. Visual system:

    10a. __172.4__ Horizontal Field-of-view
               (Average - based on 16 studies)

    10b. __199.9__ Vertical Field-of-view
               (Average - based on 15 studies)

    10c. Type visual system used:

        ___11___ Computer Generated Image (CGI)

        ___4___ Television Model Board

        ___1___ Other (describe) <u>UNSPECIFIED</u>

        ___3___ No Visual System

11. Motion system:

    Degrees-of-freedom Of Motion System: __3__ 0 (fixed base)

    __0__ 1 __0__ 2 __3__ 3 __0__ 4 __1__ 5 __12__ 6 __0__ Not Stated

    11b. Was G-seat Used? __3__ Yes __12__ No __4__ Not Stated

    11c. Was G-suit Used? __1__ Yes __4__ No __14__ Not Stated

12. Training features:

Simulator is considered a: __15__ Whole-task Trainer

__4__ Part-task Trainer

__0__ Not Stated

TRAINING CHARACTERISTICS

13. Amount of simulator training was determined by:

__4__ Proficiency-based Criterion

__14__ Blocked Design (all subjects received an equal amount of training time)

__1__ Other (describe) __UNSPECIFIED__

__0__ Not stated

14. __8.87__ Number Training Hours (summed across tasks)
    (Average - based on 16 studies)

1J. __92.4__ Number Training Trials (summed across tasks)
    (Average - based on 10 studies)

RESEARCH MEASURES

16. Dependent measures were:

__0__ Exclusively Objective In Nature

__9__ Exclusively Subjective In Nature

__10__ A Combination of Both Objective and Subjective Measures

17. Is information explicitly stated for determining initial transfer?

__5__ Yes __14__ No

18. Is information explicitly stated for determining final transfer?

__3__ Yes __15__ No

APPENDIX E

CODE BOOK

An asterisk (*) indicates the experiment may be coded in more than one category for that item.

REPORT IDENTIFICATION

1. Report ID #: _____

Write the ID # in the space provided. The report ID # appears in the top right hand corner of the front (title) page of all reports, journal articles etc. Some reports will be coded more than once, since they may have several comparison groups (e.g., Simulator Training versus Aircraft Training and Simulator Training With Motion versus Simulator Training Without Motion).

2. Coder ID #: _____

Write the coder ID # in the space provided:
Carolyn Prince (1)
Bob Hays (2)
Eduardo Salas (3)
John Jacobs (4)

3. Date of publication: _____

Write the date in the space provided. This may be found either on the front (title) page, especially if it is a journal article, or on the "report documentation page" usually placed in the first few pages of a technical report.

4. Publication Source:

_____ Book   _____ Dissertation

_____ Journal   _____ Paper Presentation

_____ Technical Report _____ Unpublished Manuscript

_____ Other (describe) _____

Place a check mark in the appropriate source category.

SIMULATOR/AIRCRAFT INFORMATION

5. Simulator name/identification code: _____

Write down the simulator's acronym, code or both (if applicable) and its classification (OFT, CPT, or WST). This information can usually be found in at least two locations: the summary of results section of the "report documentation page" and in the method section within the body of the report.

NOTE. Simulators may be identified nominally, by an alphanumeric code, or both. In general Air Force simulators are referred to by an acronym made up from their title, such as the "Advanced Simulator for Pilot Training" or ASPT. Navy simulators, by design, have an alphanumeric code, such as the 2F87F used to train P-3 pilots. A notable exception to this is the 2F103, more commonly referred to as the "Night Carrier Landing Trainer" (NCLT). Most all simulators, regardless of service branch, can be classified as either a) an operational flight trainer (OFT), b) a cockpit procedures trainer (CPT), or c) a weapons systems trainer (WST).

6. Aircraft name/identification code: _____

Many aircraft have both a name and alphanumeric identification code. An example is the P-3 "Orion". Write both in the space provided, if both are given, or at the very least, write the ID code. This information can usually found in the summary section of the "report documentation page" and in the method section within the body of the report.

RESEARCH DESIGN AND SUBJECT CHARACTERISTICS

7. Subject assignment to groups was:

_____ Random Only   _____ Matching Only

_____ Both   _____ Neither

Place a check mark in the appropriate subject assignment category. This information is usually given in the procedures section within the body of the report.

NOTE. Random selection is not the same as random assignment. If the report states that subjects are randomly selected, and there is no additional information about assignment, place a check mark in the "neither" category.

8. Subjects were:

_____ Recent Undergraduate Pilot Training (UPT) Graduates

_____ Experienced Pilots Transitioning to New Aircraft

_____ Mixed - Having Both High And Low Experienced Pilots

_____ Other (describe) _____

Place a check in the appropriate subject category. This information is normally located in the beginning of the method section within the body of the report although it may also be mentioned in the summary of the report located on the "report documentation page".

NOTE. Be careful to read any footnotes pertaining to subjects, since they may contain important information, such as subject loss or the fact that one or more subjects had additional flight experience.

*9. Group contrasts consisted of:

_____ Simulator Training Versus Aircraft-Only Training (Control)
(SIMULATOR TRAINING vs FLY ONLY CONTROL)

_____ Simulator Training With Motion System On Versus Simulator Training
With Motion System Off
(MOTION vs NO MOTION)

_____ Full Amount Of Simulator Training Versus Limited Amount Of
Simulator Training (FULL vs LIMITED TRAINING)

_____ OTHER (describe) _____

Place a check mark in one or more appropriate group contrast category, even if there appears to be no usable data from the contrast. If there is usable data, treat the separate contrasts as separate experiments and fill out another code sheet. Note on the front page of each of the code sheets that this is the first (o second, or third, etc.) code sheet for this experiment (next to the study ID #) and circle the check mark involving the group contrast category (in #9 above) specifying which contrast the code sheet corresponds.

SIMULATOR FIDELITY CHARACTERISTICS

10. Visual system:

10a. _____ Horizontal Field-of-view

10b. _____ Vertical Field-of-view

Write the exact FOV parameters in the space provided. This information is usually found in the method section within the body of the report when describing the simulator. If the study does not explicitly state this information, but cites one or more secondary sources, write in "N/S" and the secondary source(s).

   10c. Type visual system used:

      _____ Computer Generated Image (CGI)

      _____ Television Model Board

      _____ Other (describe) _____

      _____ No Visual System

11. Motion system:

   Degrees-of-freedom Of Motion System: _____ 0 (fixed base)

   ____ 1 ____ 2 ____ 3 ____ 4 ____ 5 ____ 6 ____ Not Stated

   11b. Was G-seat Used? _____ Yes _____ No _____ Not Stated

   11c. Was G-suit Used? _____ Yes _____ No _____ Not Stated

Place a check mark next to the appropriate category item for the visual system, motion DOF, and use of G-seat and G-suit. This information is usually found in the method section within the body of the report when describing the simulator. As above, if the report does not explicitly state this information, but cites one or more secondary sources, write in "N/S" and the secondary source(s).

NOTE. If the report states that one or more of these systems were "available", but doesn't state they were used, check the "Not stated" category and make a note to contact the author(s) for this information.

12. Training features:

   Simulator is considered a: _____ Whole-task Trainer

                                _____ Part-task Trainer

                                _____ Not Stated

Place a check mark next to the appropriate simulator classification category. This information may be found one of several places; the report summary, introduction, or method section. If it is not stated in the report, but the

same simulator is classified in another report, use this information, citing the other report (with page number). If two or more reports give conflicting classifications, note this also.

TRAINING CHARACTERISTICS

13. Amount of simulator training was determined by:

_____ Proficiency-based Criterion

_____ Blocked Design (all subjects received an equal amount of training time)

_____ Other (describe) _____

_____ Not stated

Place a check mark next to the appropriate training category. This information is usually found in the procedure section of the body of the report when describing training procedures for the experimental and control group. In some cases, performance outcomes measures are trials-to-proficiency, but training wasn't stopped once proficient performance was reached. Thus, a category other than proficiency training should be checked reflecting the actual training procedure.

14. _____ Number Training Hours (total)

15. _____ Number Training Trials (total)

Write number(s) in space provided. Both training hours and trials may not be given within the report. If one or both are not given, write "N/S". This information can usually found in the procedure section within the body of the report and/or given in a table specifying the training syllabus used. If conflicting information is given by two or more sources within the report, note both with accompanying page numbers where information is found.

NOTE. The term trial should be understood to mean a single repetition of a given task or set of tasks. Trials should not be confused with sessions on the simulator or sorties in the aircraft, since multiple trials may occur within a given session/sortie.

RESEARCH MEASURES

16. Dependent measures were:

_____ Exclusively Objective In Nature

_____ Exclusively Subjective In Nature

_____ A Combination of Both Objective and Subjective Measures

Place a check mark next to the appropriate measure category. Dependent measures are usually discussed in the methods section within the body of the report. Some reports include additional information about the dependent measure(s) in an appendix. If present, read this information carefully, since it may prove critical for determining the correct classification of the measure(s).

NOTE. Trials-to-proficiency (criteria), when proficiency is determined by subjective IP ratings should be classified as subjective.

17. Is information explicitly stated for determining initial transfer?

_____ Yes _____ No

18. Is information explicitly stated for determining final transfer?

_____ Yes _____ No

Place a check mark in the appropriate response category. Information about initial and final transfer can usually be found either in the method section within the body of the report, in a table summarizing transfer-of-training performance, or in an appendix. In some instances, multiple trials (e.g., the first through the third) are used to assess initial transfer and later trials (e.g., the seventh and eighth) are used to assess final transfer. Note the transfer trial used for both of these measures when the information is present.

APPENDIX F

PROCEDURES FOR CALCULATING WEIGHTED MEAN
POINT BISERAL CORRELATION COEFFICIENTS

Procedures for coming up with a single training outcome effect size (ES) estimate for individual experiments were as follows. Experiments reporting usable study statistics (i.e., values based on t-tests, F-tests, Chi Squares, and Mann-Whitney U tests) were converted to point biserial correlation coefficients using formulas provided by Glass et al. (1981; 1970). When appropriate study statistics were omitted, a search for information allowing calculation of usable statistics was performed. In many such cases, means and standard deviations were obtained and subsequently used to calculate one or more t- statistics. If the number of subjects (N's) for corresponding experimental and control groups were disparate, the resulting t- value was corrected using a formula described by Hunter et al. (1982, p. 99). In other cases, reported raw data were used to calculate an F-statistic or Chi Square. In four experiments, information was reported describing the size of a treatment effect using only a p value (with associated treatment means). In order to render this value usable, it was first transformed to a t-value based on corresponding degrees of freedom and conservative alpha level (using a one-tailed distribution) and subsequently transformed into a point biserial correlation coefficient. In all cases, whenever more than one research statistic was reported in a single experiment, an average point biserial correlation coefficient was calculated by first weighting individual research statistics by the number of subjects used to calculate the statistic. The final, weighted mean correlation coefficient, denoted as RPB, for a given experiment has an attached weight equal to the mean number of subjects used to calculate the individual research statistics. This weight is used when calculating the overall (population) effect size estimate across experiments.

In cases where there existed competing values that could be used to estimate a training outcome effect size, the most conservative value was chosen. For example, when converting reported p values to their corresponding t-value, the one-tailed table value was used since this value is smaller, thereby providing a more conservative effect size estimate than the two-tailed value. The exception to this was converting Mann-Whitney U values to corresponding t-values. Glass et al. (1981, pp. 130-131) note that a U-statistic at the .05 probability level corresponds to a t-statistic at the .03 or .02 level and thereby provides a more conservative effect size estimate than the standard t-test. Since the U-statistic is more conservative than a corresponding t-statistic (assuming the same alpha level), use of the two-tailed value when converting the former to the latter is appropriate, as opposed to a smaller one-tailed value.

Note 1. Conversion of t- and F-values were done using a FORTRAN- based program run on a Zenith 248 microcomputer. This "metatran" program was generously supplied by Dr. Terry Dickinson of Old Dominion University Department of Psychology.

APPENDIX G

SUBGROUP ANALYSIS COMPARING JET AND HELICOPTER EXPERIMENTS

Aircraft Type	Number of articles	$\overline{RPB}$	S^2_{RPB}	S^2_e	S^2_T	% un-explained variance
Jets	10	.26	.03119	.01979	.01140	37
Helicopters	3	.02	.00201	.01786	-	0
Total Group	13	.19	.03452	.01920	.01532	44

Note. $\overline{RPB}$ is weighted mean point biserial correlation coefficient. S^2_{RPB} is observed variance. S^2_e is error variance. S^2_T is "true" variance. The dash (-) should be read as a zero.

APPENDIX H

SUBGROUP ANALYSIS COMPARING THREE TYPES OF JET EXPERIMENTS

Experiment Type	Number of articles	$\overline{RPB}$	S^2_{RPB}	S^2_e	S^2_T	% un-explained variance
Simulator versus Aircraft-Only Training	10	.26	.03119	.01979	.01140	37
Motion versus No-Motion	5	-.05	.03673	.03706	-	0
Other	4	.19	.03559	.01303	.02256	63
Total Group	19	.15	.04256	.02143	.02113	50

Note. $\overline{RPB}$ is weighted mean point biserial correlation coefficient. S^2_{RPB} is observed variance. S^2_e is error variance. S^2_T is "true" variance. The dash (-) should be read as a zero.

APPENDIX I

CONFIDENCE INTERVALS AND RELATED VALUES FOR $\overline{RPB}$
TABLE VALUES BY AIRCRAFT AND EXPERIMENT TYPE

Result Classification Category by Aircraft and Experiment Type	$\overline{RPB}$ Value	Sd	95% CI Values (- to +)
JETS			
Simulator versus Aircraft-Only Training			
(1) Overall effect size	.26	.014	.23 to .29
(2) Objective measures only	.12	.061	.00 to .24
(3) Subjective measures only	.25	.015	.22 to .28
(4) Initial transfer trial	.19	.032	.13 to .25
(5) Final transfer trial	.03	.224	-.42 to .48
Motion versus No Motion			
(1) Overall effect size	-.05	.057	-.16 to .06
HELICOPTERS			
Simulator versus Aircraft-Only Training			
(1) Overall effect size	.02	.019	-.02 to .06

Note. $\overline{RPB}$ value is mean weighted point biserial correlation coefficient. Sd is the associated standard deviation for a given RPB value. CI means "confidence interval".

APPENDIX J

SUBGROUP ANALYSIS OF RESEARCH CHARACTERISTICS FOR JET EXPERIMENTS

Moderator (by subset)	Number of articles	$\overline{RPB}$	S^2_{RPB}	S^2_e	S^2_T	% un-explained variance
Simulator Fidelity Characteristics						
Did the visual system incorporate computer generated imaging (CGI) technology?						
Yes[a]	5	.21	.02768	.01720	.01049	38
No	2	.50	.01139	.01785	-	0
What type trainer was simulator:						
Part-task[b]	2	.12	.00010	.01354	-	0
Whole-task	8	.35	.02890	.02355	.00535	19
Training Characteristics						
What type training system was employed?						
Blocked	6	.21	.02076	.01972	.00103	5
Proficiency	3	.54	.00623	.02461	-	0
Research Measures						
Did the dependent measures employed in the experiment include both objective and subjective measures?						
Yes	3	.12	.00010	.01631	-	0
No	7	.39	.02376	.02285	.00091	4
Total Group	10	.26	.03119	.01979	.01140	37

[a]Item eliminated since it failed to demonstrate a reduction in true variance relative to that of the total group.

[b]Variables incorporated into only two (2) experiments were reported soley for the purpose of visual inspection and should not be considered a valid moderator based on this analysis.

Note: $\overline{RPB}$ is weighted mean point biserial correlation coefficient. S^2_{RPB} is observed variance. S^2_e is error variance. S^2_T is "true" variance A dash (-) should be read as a zero (0).

APPENDIX K

INTERCORRELATIONS AMONG RESEARCH CHARACTERISTICS
FOR JET EXPERIMENTS

RESEARCH CHARACTERISTICS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1) Use of matching prior to subject assignment	1.00									
2) Use of CGI visual system	-.548 (7) p=.102	1.00								
3) Total FOV of visual system (C)	-.212 (9) p=.292	.588 (7) p=.083	1.00							
4) DOF of motion system (C)	.110 (10) p=.381	-.258 (7) p=.288	.355 (9) p=.174	1.00						
5) Use of G-seat	.293 (8) p=.241	N/A	N/A	.151 (8) p=.360	1.00					
6) Use of whole-task simulator	.218 (10) p=.272	-.258 (7) p=.288	.245 (9) p=.262	.926* (10) p=.000	.218 (8) p=.302	1.00				
7) Use of proficiency-based training	.500 (9) p=.085	-.300 (7) p=.257	-.133 (8) p=.377	.320 (9) p=.201	.646 (7) p=.059	.378 (9) p=.158	1.00			
8) Use of objective and subjective dependent measures	-.524 (10) p=.060	.400 (7) p=.187	.189 (9) p=.313	-.662* (10) p=.019	-.293 (8) p=.241	-.764* (10) p=.005	-.500 (9) p=.085	1.00		
9) Number training hours (C)	.369 (9) p=.164	.008 (6) p=.494	.244 (8) p=.280	.476 (9) p=.098	-.152 (7) p=.373	.412 (9) p=.135	.532 (8) p=.087	-.447 (9) p=.114	1.00	
10) Number training trials (C)	.296 (5) p=.314	.455 (5) p=.220	.943* (5) p=.008	.338 (5) p=.289	N/A	.338 (5) p=.289	N/A	.008 (5) p=.495	.381 (4) p=.309	1.00

Notes: Number in parentheses is number of experiments used when calculating Pearson correlation coefficient. N/A indicates correlation coefficient can not be calculated. * indicates correlation coefficient has $p<.05$. (C) indicates variable is continuous.

APPENDIX L

SUBGROUP ANALYSIS COMPARING THREE TYPES OF JET TASKS

Task Type	Number of articles	$\overline{RPB}$	S^2_{RPB}	S^2_e	S^2_T	% un-explained variance
Takeoff	3	.65	.02849	.01438	.01410	50
Approach	3	.64	.00646	.01273	.	0
Landing	3	.57	.03081	.01695	.01385	45
Total Group	10	.26	.03199	.01979	.01140	37

Note: $\overline{RPB}$ is weighted mean point biserial correlation coefficient. S^2_{RPB} is observed variance. S^2_e is error variance. S^2_T is "true" variance.

APPENDIX M

TASK DIFFICULTY SURVEY FOR JETS TASKS

BACKGROUND INFORMATION

RANK:_____   EVER BEEN AN INSTRUCTOR PILOT?   YES   NO
                                   (circle one)

YEARS FLYING:_____

TYPES OF AIRCRAFT FLOWN:_____
                    (list in order of hours of experience - most to least)

TYPES OF SIMULATORS TRAINED ON:(list)_____

DIRECTIONS - On the following pages are listed several maneuvers/tasks often trained using a simulator. Use the 1-3 scale below to rate each task in terms of **how difficult the task is to learn**. If you first learned the task on a simulator, rate how difficult the task was to learn **while training on the simulator** (as opposed to aircraft). Then circle the item number corresponding to the simulator trained task (see the example on the top of the next page).

Place a rating number on the line next to each task. Place a ZERO (0) next to any task that you are unsure or haven't performed.

1	2	3
. . . .		
LOW	MEDIUM	HIGH
DIFFICULTY	DIFFICULTY	DIFFICULTY

1. A LOW DIFFICULTY task is one in which:
     -actions are clearly defined
     -all information is available
     -components of task can be learned
       in a short period of time

   2. A MEDIUM DIFFICULTY task is one which:
         -it is not always clear what your actions should be
         -needed information may not always be available
         -performance of the task is often a series of
           actions that are moderately complex
         -there is some stress involved

3. A HIGH DIFFICULTY task is one which:
- there are a number of things to do
- needed information may not be present
- adaptation is required
- stress is moderate to high
- actions that make up task are moderately to very complex

EXAMPLE: (helicopter tasks)

__2__ 1. Takeoff to Hover (simulator trained task)
__1__ 2. Landing from Hover
__3__ 3. Confined Area Approach (simulator trained task)

General Description: Air-to-air combat maneuvers

<u>Maneuver/task description</u>

(Related Skills)

_____ 1. Acceleration Maneuver   _____ A. Descriptive Commentary
_____ 2. High Yo-Yo   _____ B. Range Estimation
_____ 3. Quarter Plane   _____ C. Target Acquisition
_____ 4. Barrel Roll Attack   _____ D. Kept Bogey in Sight
_____ 5. Immelmann Attack   _____ E. Weapons Parameter
_____ 6. Lag Roll   Recognition
_____ 7. Separation   _____ F. Switchology
_____ 8. Tactical Formation
_____ 9. Set up on Perch
_____10. Defensive Maneuvers
_____11. Low Yo-Yo
_____12. Lag Pursuit
_____13. Rolling Scissors
_____14. Guns Defense (High-G Barrel Rolls)
_____15. Head On Maneuvering
_____16. Atoll Extension

General description: Four engine jets only (if you have never flown this type aircraft, please skip this section)

<u>Maneuver/task description</u>

_____17. Abort Four Engines
_____18. Abort Three Engines
_____19. Engine Failure After Refusal
_____20. Departure
_____21. Holding

_____22. TACAN/VOR
_____23. LOC
_____24. GCA
_____25. ILS
_____26. Normal Landings
_____27. Approach Flap Landings
_____28. Waveoff
_____29. Three Engine Landings

General Description: Airwork maneuvers

Maneuver/task description

_____30. One Engine Failure at TO
_____31. Two Engine Failure at TO
_____32. Low Altitude Restart
_____33. One Engine Approach
_____34. Hydraulic Failure
_____35. Slow Flight
_____36. Takeoff
_____37. Straight in Touch-and-go
_____38. Go Round
_____39. TO Climb
_____40. Landing
_____41. Traffic Pattern Stall
_____42. Control Response
_____43. Trim
_____44. Straight-and-level
_____45. Pitch, Bank, and Power
_____46. Constant Air Speed (CAS) Straight-and-level
_____47. CAS Climb
_____48. CAS Descent
_____49. CAS Climb Turn
_____50. Level Offs
_____51. Level Turns
_____52. Change of Airspeed
_____53. Traffic Pattern Steep Turns
_____54. 30 deg. Bank Turns
_____55. 45 deg. Bank Turns
_____56. 60 deg. Bank Turns
_____57. Turn-to-headings (TH)
_____58. Airspeed Changes While TH
_____59. Tech Order Climbs
_____60. Configuration Change
_____61. 30 deg. Bank Descending Left Turns
_____62. Traffic Exits
_____63. Straight in Approach Landings
_____64. 360 deg. Traffic Pattern
_____65. Power-on Stalls
_____66. Constant Air Speed (CAS) Descending Turn

_____67. Vertical-S-Delta
_____68. Carrier Qualification (CQ) Landings
_____69. Night CQ Landings
_____70. Field Carrier Landing Practice (FCLP)
_____71. Night FCLP
_____72. Bomb Delivery Approach
_____73. Bomb Delivery Release
_____74. Air-to-air Refueling

General Description: Formation Flying

Maneuver/task description

_____75. Fingertip
_____76. Crossunder
_____77. Turning Rejoin
_____78. Wingwork (Fingertip at 15-30 deg. bank)
_____79. Procedures (start up & shut down)
_____80. Aborted TO

General Description: Aerobatics

Maneuver/task description

_____81. Aileron Roll
_____82. Split S
_____83. Loop
_____84. Lazy 8
_____85. Immelmann
_____86. Cuban 8
_____87. Cloverleaf

DIRECTIONS: Now go back and for tasks having prior simulator training (circled items), place a plus (+) next to the rating if actual performance in aircraft was noticeably harder than in simulator. Place a dash (--) if performance in aircraft was noticeably easier than in simulator (see example below).

EXAMPLE:

_-2_ 1. Takeoff to Hover (simulator trained task that is easier in aircraft)

_1_ 2. Landing from Hover

_+3_ 3. Confined area Approach (simulator trained task that is harder in aircraft)

APPENDIX N

GUIDELINES FOR REPORTING FUTURE EXPERIMENTAL RESULTS

Introduction

The problems associated with conducting transfer-of-training (TOT) research in the aviation training domain are well documented. It is fully expected that future research in this area will be plagued by similar problems and that experimental rigor will suffer accordingly. These guidelines are meant as an aid for those attempting TOT experiments so that sufficient information will be available for subsequent meta-analytic review. While meta-analysis has its own set of problems, it offers a unique perspective for answering many long-standing questions about the nature of simulator training.

In general, consider that any information that is not explicitly stated within the report can not be assumed by the reviewer. For example, several reports included in this review noted that subjects were randomly selected from a class of student aviators. However, no mention was made concerning student assignment to the experimental conditions. The following items are sources of information that should be addressed when reporting results of experimentation in this area.

1) Research design

a) Matching - State whether subjects were matched prior to assignment to the experimental conditions. Describe the variable(s) used for matching and the outcome of the matching procedure.

b) Subject assignment - State the procedure for assigning subjects to the experimental conditions. If random assignment wasn't possible or was compromised in any way, report information about how it affected the various groups.

c) Loss of subjects - Attrition may occur for a variety of reasons and information concerning subject loss must be described in detail, including procedures used when performing statistical analyses (e.g., adjusting degrees- of-freedom).

d) Bias reduction procedures - Counterbalancing and having raters blind to subject's experimental assignment are common procedures used to reduce potential measurement bias. Their use (or non-use) should be chronicled.

e) Estimation of rater agreement - Inter- and (if possible) intra-rater agreement should be assessed and reported. If objective measures are used in conjunction with with subjective indices, report the relation (e.g., r) between these measurement types.

2) <u>Performance measures and statistical analysis</u>

    a) Performance measures - A detailed description of each performance measure should be given. If an established measure is used, cite relevant literature describing each measure and give pertinent information concerning its application within the experiment.

    b) General statistical reporting requirements - Reporting detailed information about all statistical analyses is a must. In general, means, associated standard deviations, and number of subjects must be reported for separate analyses, even when multivariate procedures are used.

    c) Commonly used statistical procedures - Common statistical procedures used to describe the magnitude of between-group differences include t- and F-tests. The value of the statistic should be reported as well as the associated p-value. An indication of the direction of the statistic, relative to the stated null hypothesis, should also be stated, and apriori versus aposteriori analyses should be delineated. The same basic requirements apply when reporting analyses based on non-parametric procedures (e.g., chi square). Reporting exact p-values is imperative when reporting results of non-parametric analyses.

    d) Areas of specific interest - Currently, issues surrounding task-type and the extent of TOT performance are in need of investigation. Accordingly, information about individual tasks should be reported separately as well as that for individual transfer trials (if appropriate).

    e) Covariates - Use of any cognitive or non-cognitive variables as covariates should be reported. The relationship (r) between the covariate and associated criterion performance variable should be given, and if appropriate, the interclass correlations between the various criterion measures.

3) <u>Training characteristics</u>

    a) General training features - Describe the extent to which the various training programs used in the experiment were alike and how they differed in terms of relevant training parameters (e.g., time to complete training, number of training trials).

    b) Instructor variables - Report the instructor-student ratio as well as the use of specific simulator instructional features by the instructor. Describe any training given to instructors on the use of instructional features. Describe the instructor's level of acceptance of the simulator as a training device (this information may be task-specific). Describe the extent to which student motivation influences instructor ratings.

c) Student variables - Report the student's level of acceptance of the simulator as a training device. Describe any attempts at transitioning student from simulator to aircraft.

d) Training program - If applicable, describe the development of the training program(s) with regard to ISD principles and procedures or at least cite references providing this information. Describe how knowledge-of-results (KOR) was given to the student (e.g., how often, in what form). Describe any part-task training methods employed.

4) **Simulator fidelity characteristics**

a) General - Describe the level of physical and functional fidelity for each of the simulator's subsystems (i.e., sound, motion/force, visual, cockpit display, and flight control characteristics). The key here is reporting use, and not just availability, of specific simulator components (e.g., g-seat, g-suit, stick shaker system).

b) Specific areas of interest - Describe how the simulator flight control characteristics were validated (example: seat-of-pants vs. data from actual aircraft). If use of motion/force cuing is a primary experimental manipulation, provide information concerning calibration of hardware/software parameters (at very least, report results of calibration tests prior to and after experimentation).

DISTRIBUTION LIST

Commanding Officer
Naval Aerospace Medical Institute
Naval Air Station
Code OOL
Pensacola, FL 32508-5600

Commander
Naval Air Systems Command
Naval Air Systems Command Headquarters
AIR 950D - Technical Library
Washington, DC 20361-0001

Commander
Naval Air Systems Command
Naval Air Systems Command Headquareters
Code 933G
ATTN: CDR Steve Harris
Washington, DC 20361-0001

Commander
Naval Air Systems Command
Naval Air Systems Command Headquarters
Code PMA 205
ATTN: LCDR Mike Lilienthal
Washington, DC 20361-0001

Commander
Naval Research Laboratory
Library
Washington, DC 20375

Mr. B. G. Williams
Chief of Naval Education and Training
Naval Training Systems Center
Liaison Office, Code L02
Bldg 268, Room 2-44
NAS, Pensacola, FL 32500-5100

Chief of Naval Education and Training
Code N6
NAS, Pensacola, FL 32508-5100

Commanding Officer
NETPMSA
Code 04
ATTN: Dr. William Maloy
NAS Pensacola, FL 32509-5100

Commanding Officer
Naval Training Systems Center
Marine Corps Liaison Office
Code 001
Orlando, FL 32826-3224

Commanding Officer
Naval Training Systems Center
Air Force Liaison Office
Code 002
Orlando, FL 32826-3224

Commanding Officer
Naval Training Systems Center
Code 1
Orlando, FL 32826-3224

Commanding Officer
Naval Training Systems Center
Code 2
Orlando, FL 32826-3224

Commanding Officer
Naval Training Systems Center
Code 25
Orlando, FL 32826-3224

Commanding Officer
Naval Training Systems Center
Code 26
Orlando, FL 32826-3224

Commanding Officer
Naval Training Systems Center
Code 35
Orlando, FL 32826-3224

Commanding Officer
Navy Personnel Research and Development Center
Code 01
ATTN: Dr. R.C. Sorenson
San Diego, CA 92152-6800

Commanding Officer
Navy Personnel Research and Development Center
Code 51
ATTN: Dr. W. Wulfeck
San Diego, CA 92152-6800

Commanding Officer
Navy Personnel Research and Development Center
Code 52
ATTN: Dr. J. McLachlan
San Diego, CA 92152-6800

Department of the Navy
Office of the Chief of Naval Operations
OP-01B2
Washington, DC 20350-2000

Department of the Navy
Office of the Chief of Naval Operations
OP-987H
ATTN: Dr. Bart Kuhn
Washington, DC 20350-2000

Department of the Navy
Office of the Chief of Naval Operations
OP-01B2E1
Washington, DC 20350-2000

Chief of Naval Research
Office of the Chief of Naval Research
Code 1142PT
ATTN: Dr. Susan Chipman
Arlington, VA 22217-5000

Chief of Naval Research
Office of the Chief of Naval Research
Code 1142 EP
Arlington, VA 22217-5000

Office of Naval Technology
Code 222
ATTN: Dr. Stan Collyer
800 North Quincy Street
Arlington, VA 22217-5000

Chief, U.S. Army Research Institute
Orlando Field Unit
ATTN: Dr. Stephen Goldberg
12350 Research Parkway
Orlando, FL 32826

Chief, U.S. Army Research Institute
Field Unit
PERI-IN
Ft. Rucker, AL 36362-5354

PMTRADE
AMPCM-TND-E
ATTN: Dr. R. Hofer
Orlando, FL 32826-3276

Chief, U.S. Army Research Institute
ATTN: Dr. J. Hiller
5001 Eisenhower Avenue
Alexandria, VA 22333

AFAMRL/HEF
Director, Human Egnineering Division
Wright-Patterson AFB, OH 45433

AFHRL
Training Systems
ATTN: Col W.J. Lobbestael
Wright-Patterson AFB, OH 45433

AFHRL
Training Systems
ATTN: Col Michael Lane
Williams AFB, AZ 85240

AFHRL/OTT
ATTN: Dr. Dee Andrews
Williams AFB, AZ 85240-6457

AFAMRL/HEC
ATTN: Dr. John Forester
Bldg 248, Area B
Wright Patterson AFB, OH 45433-6583

Defense Technical Information Center
FDAC
ATTN: J. E. Cundiff
Cameron Station
Alexandria, VA 22304-6145

Marine Corps Program Manager
for Training Systems
Code SST
Marine Corps Research, Development
and Acquisition Command
Quantico, VA   22134

Institute for Defense Analyses
Science and Technology Division
ATTN: Dr. Jesse Orlansky
1801 Beauregard Street
Arlington, VA   22311

National Defense Institute
Research Directorate
Ft McNair, DC   20319

Director, TPDC
3280 Progress Drive
Orlando, FL   32826

DASD (FM&P/Tng Policy)
ATTN: Mr. Gary Boycan
The Pentagon, Rm 3B930
Washington, DC   20301-4000

Institute for Simulation and Training
ATTN: Dr. L. Medin
12424 Research Parkway, Suite 300
Orlando, FL   32826