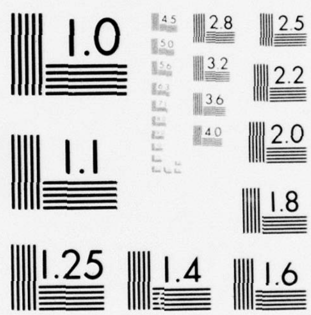
AD-A036 460 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OHIO SCH--ETC F/G 14/2
TECHNIQUES FOR THE INITIAL EVALUATION OF FLIGHT SIMULATOR EFFEC--ETC(U)
DEC 76 R L MILLER

UNCLASSIFIED GSM/SM/76D-34 NL

1 OF 2
ADA036460



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

ADA036460

DDC
RECEIVED
MAR 7 1977
C

UNITED STATES AIR FORCE

AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

D D C
MAR 7 1977
C

TECHNIQUES FOR THE INITIAL EVALUATION OF FLIGHT SIMULATOR EFFECTIVENESS

THESIS

GSM/SM/76D-34

Ralph L. Miller
Capt USAF

Approved for public release; distribution unlimited.

GSM/SM/76D-34

TECHNIQUES FOR THE INITIAL EVALUATION OF FLIGHT SIMULATOR EFFECTIVENESS

THESIS

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

ACCESSION for
NTIS White Section
DDC Buff Section
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
Dist. AVAIL. and/or SPECIAL
A

by

Ralph L. Miller, B.S.

Captain USAF

Graduate Systems Management

December 1976

Approved for public release; distribution unlimited.

PREFACE

The increasing cost and complexity of modern flight simulators has caused development and evaluation of these devices to become very complex tasks. This thesis represents my perceptions of these tasks and one technique for accomplishing them. The subject of flight simulator development and evaluation was first brought to my attention by Major James A. Richmond of the Simulator System Program Office at Wright-Patterson Air Force Base, Ohio. Major Richmond has offered his enthusiastic support throughout this research effort. Without his help, this research would not have been possible.

I am particularly grateful to Major G. C. Saul Young, my advisor, and Major Charles W. McNichols, my reader. Their guidance and understanding contributed immeasurably to the completion of this project. I would also like to express my appreciation to Miss Joyce L. Wilson who typed this report under the pressures of a very tight schedule. Finally, I would like to acknowledge the understanding, endurance, and cooperation of my wife, Lea, who kept the home first burning during the period of this research effort.

Ralph L. Miller

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

This report presents an analysis of the development and initial evaluation of Air Force flight simulators. The objectives of the study were to determine the criterion variables most applicable to an initial flight simulator evaluation and to develop a general technique for the evaluation of these criterion variables.

The research began with a review of current Navy, Army, and Air Force flight simulator development and evaluation techniques. This review, combined with information gathered from related sources, provided the basis for examination and selection of criterion variables. The variables examined by this effort were: aircraft flight time saved, training efficiency, transfer of training, fidelity of psychological simulation, fidelity of engineering simulation, and simulator effectiveness. The examination of these variables concentrated on their measurability during an initial flight simulator evaluation and their ability to predict how well a flight simulator would perform its intended mission.

Following the examination of criterion variables, the research concentrated on the development of a technique for the evaluation of applicable criterion variables. The resulting technique is a combination of the traditional quantitative techniques plus some subjective techniques. The purpose of the subjective techniques is to identify simulator characteristics that are perceived to be different from the real work aircraft characteristics and to assess the impact that these differences will have on the operational use of the flight simulator.

TECHNIQUES FOR THE INITIAL EVALUATION OF FLIGHT SIMULATOR EFFECTIVENESS

I. The Research Problem

Introduction

The sophistication and cost of flight simulation has increased dramatically during the last century. Although all of the ground training devices used to teach flying skills are usually called flight simulators, two main classifications exist. These are procedural trainers and flight simulators.

Procedural trainers were the only form of ground training device in use until the early 1960's. These trainers usually represented an entire class or type of aircraft, unlike modern flight simulators which are representative of only one specific aircraft (Williges, et al., 1972). The other major difference is the ability of the device to duplicate the aircraft handling characteristics or control feel. Flight simulators reproduce the feel of the controls throughout the flight envelope of the aircraft represented. Procedural trainers do not duplicate control feel and are used to teach procedural steps rather than handling characteristics. Most ground training devices currently used by the military are procedural trainers. Almost all new training devices being purchased would be properly classified as flight simulators.

The increasing cost and complexity of modern flight simulators has required Test and Evaluation (T&E) to become one of the key functions performed during the acquisition of an Air Force flight simulator. The T&E procedures used must insure that the flight simulator being purchased will be capable of effectively and efficiently accomplishing the training mission for which it was designed and purchased.

The increased complexity of modern flight simulators has also made the evaluation of effectiveness extremely difficult. Very little is known about the factors which influence training effectiveness or how these factors must interact to produce an effective flight simulator. The purpose of this report is to determine the criterion variables for flight simulator evaluation and to develop a technique for an initial evaluation of the simulator effectiveness using these criterion variables.

Background of Flight Simulation

History has shown that military aviation has usually been slower to accept and use ground based flight simulation than has the civilian aviation community. At the close of World War I, flight simulation for the military consisted of a limited number of short-winged training airplanes. These planes could only be taxied on the ground because they were incapable of flight. Aviation students ran them up and down the flying field in order to learn the concept of working aircraft controls prior to the first actual aircraft flight. These trainers became known as "Stub-winged Jennies" or "Grass-cutters" (Williges, et al., 1972).

During the same time period, civilian aviators were using fixed-base flight simulators that had been developed from the "Sanders Teacher" and "Eardly-Billing Oscillator" of the early 1900s. These trainers consisted of an aircraft mock-up mounted on a single pivot point which allowed limited movement in pitch, yaw, and roll. A mechanical linkage connected to the flight controls provided the appropriate movement for any control input (Lewis, 1974). Although these trainers had the same purpose as the "Stub-winged Jennies," they provided more efficient as well as more effective training, in that they allowed visualization of aircraft response to control inputs.

The military aviation attitudes about flight simulation became favorable for the first time during the late 1920s with the introduction of "blind" or instrument flying. This facet of aviation training proved to be excessively dangerous and uneconomical to conduct in the aircraft (Lewis, 1974). Both military and civilian aviation realized that a safer and more efficient means of training was needed. The military became an interested observer as civilian simulator manufacturers attempted to develop an instrument flight simulator with the existing mechanical linkage technology.

Instrument flight simulation was first accomplished in 1929 by a newcomer to the flight simulation industry. Edwin A. Link was able to combine his intense personal aviation interest with the experience and knowledge he had gained while working at the Link Piano and Organ Company to produce an effective instrument flight simulator. His simulator used a system of motors, bellows, and mechanical linkage from the organ factory to produce the illusion of instrument flight (Heinle, 1973). However, Mr. Link was not able to gain the interest of military or civilian aviation with his first simulator. It sold almost exclusively to amusement parks for from $300 to $500 (Snow, 1975).

Mr. Link continued to improve his simulator and finally sold the first military flight simulator to the Navy in 1931 for $1,500. The evaluation of this first military simulator took the form of a demonstration of effectiveness. The Navy became convinced that the flight simulator was effective after Mr. Link taught an officer who had never been in an aircraft to fly by instruments in his device (Snow, 1975).

The first instrument flight simulator used extensively in the military was the Link model C-3, better known as the "Blue Box." This simulator used a pneumatic computation system to simulate aircraft motion and instrument readings. Thousands of military pilots were trained in this device during World War II. The Link "Blue Box" and the simulators that preceded it fit into the procedural trainer rather than the flight simulator classification. The purpose of these early trainers was to teach the concept of using controls and instruments. No attempt had been made to duplicate the handling characteristics or "feel" of the aircraft being simulated (Heinle, 1973).

Aviation technology advanced rapidly following World War II. Electrically driven flight instruments began replacing the older mechanical instruments during the late 1940's (Dunlap, et al., 1975). One of the most important developments of this time period was the jet engine Jet aircraft provided much higher performance and were far more complex than aircraft previously used. The flight simulators developed to train pilots in these aircraft also accelerated in sophistication (Rhodes, 1967). The complexity of high performance jet aircraft simulation required more advanced computational capabilities than were available with the current technology.

Electronic computers were developed and began to be applied to flight simulation during the 1950's. The first generation of computers consisted of alternating current carrier analog computers. Simulation with these computers was limited and characterized by inaccurate computation, poor reliability, limited capacity, large space requirements, and a dependence on very scarce, highly skilled analog computer programmers to avoid build-up of computational errors (Fogarity, 1967).

The vast difference between the high performance characteristics of jet aircraft and those of previous aircraft caused extensive training problems for the military. Military aviation hoped to solve these training problems by purchasing flight simulators that accurately simulated handling characteristics through the use of the newly developed electronic computers. The Link C-11A simulator was built in 1950 for this purpose. Simulation of aircraft motion and handling characteristics proved to be impractical due to the static accuracy and dynamic response limitations of the analog computers in use at that time (Fogarity, 1967).

The increasing cost and complexity of electronic simulation compounded by the inability to accurately reproduce aircraft handling characteristics caused military aviation to seriously question the training value of flight simulation. The military believed the potential of flight simulation was limited to teaching basic cockpit, instrument, and emergency procedures and that the only way to teach handling characteristics was in the aircraft. This attitude prevailed through the 1960's (Lewis, 1974).

The introduction of real time medium sized digital computers in the early 1960's provided the computational technology necessary for complete simulation of all the sensations of flight (Fogarity, 1967). The large capacity, high speed, and accuracy of computation in these computers made it possible to duplicate aircraft handling characteristics as well as incorporate motion and visual displays with instrument flight simulation (Heinle, 1973).

Major political and economic events of the 1970's forced the perceived value and purpose of flight simulation in the military to change again. The United States involvement in Southeast Asia resulted in large military budgets which had become politically unpopular. The termination of this involvement allowed Congress to drastically reduce the military portion of the federal budget. The implementation of an all volunteer military force during this same time period required a much larger portion of the remaining military budget to be devoted to personnel costs. These two events caused a squeeze on the defense dollars available for operations and aircraft procurement. This squeeze was compounded by high fuel prices resulting from the 1973 oil embargo and escalating costs of aircraft procurement due to inflation (Dunlap, et al., 1975).

These political and economic forces caused the United States Air Force to undertake several studies to determine how effective flight simulators would be as a means to reduce operating costs and prolong the life of existing operational aircraft. These studies concluded that the increased use of flight simulators could effectively reduce operational costs. This conclusion was based on the fact that most of the peacetime flying in the Air Force is devoted to training (Dunlap, et al., 1975). The validity of this conclusion was dependent on the avility of flight simulator sorties to be substituted for aircraft training sorties without significantly reducing the quality of train training.

Statement of the Problem

Introduction to the Problem. This research effort will concentrate on the flight simulator evaluation problems resulting from two of the major events in the recent history of flight simulation. These events were: (1) the introduction of digital computers during the 1960's which provided the capability to build flight simulators which could duplicate nearly all real world aircraft characteristics, and (2) the political and economic forces during the early 1970's which compelled the Air Force to place substantially more emphasis on the use of flight simulators.

The improved capability of digital computer flight simulation turned both the design and evaluation of flight simulators into very complex and difficult tasks. In order to program the computers for detailed and complete simulation, an accurate mathematical model of the aircraft characteristics had to be constructed (Dunlap, et al., 1975). Six equations, one for each degree of freedom (pitch, roll, yaw, vertical, lateral, and longitudinal), were required to define motion alone (Catron, 1975). The model also had to include equations for instrumentation, visual display, control forces, aircraft systems, etc. The data required for these equations has normally been derived from preproduction engineering design studies (Dunlap, et al., 1975). However, refining the mathematical representation of the aircraft is not a primary goal of the airframe manufacturer. Therefore, the definitive data required to completely describe the aircraft characteristics under all flying conditions and throughout all flight regimes was simply not available (Catron, 1975).

Since the flight simulators had to be designed with incomplete and inaccurate data, the traditional use of aircraft flight data as the standard of performance during the evaluation became untenable. Evaluation relative to incomplete data only measured how well the simulator was programmed and gave no indication of how accurately the real world aircraft had been simulated or how effective the flight simulator would be. With traditional evaluation techniques degraded, a new evaluation technique became desperately needed.

The political and economic events of the 1970's added new requirements to Air Force flight simulation. In order to resolve the problems attributable to these events, new flight simulators were required to be capable of replacing some of the training traditionally conducted in the aircraft as well as be an effective medium for teaching new skills. The extent to which flight simulators can counteract the defense budget squeeze, the energy crisis, the escalating cost of aircraft procurement and operations, and the need to extend the life of current operational aircraft must be considered in developing any new evaluation technique (Williges, et al., 1972).

Before an evaluation technique can be developed, flight simulator properties must be examined. The relationship between each of these properties and the eventual effectiveness of the flight simulator must also be known. If the relationship between properties and effectiveness is well defined, an evalution technique which measures one or more of the flight simulator properties may be feasible. This approach is very difficult since few of the properties of flight simulators are well defined and the relationship of individual properties to flight simulator effectiveness is frequently unknown. The difficulty is compounded by the fact that many of the simulator properties cannot be accurately measured.

The Problem. In order to develop a technique for the initial evaluation of flight simulator effectiveness, two questions must be answered:

1. Which flight simulator properties should be used as criterion variables for evaluation in order to estimate flight simulator effectiveness?

2. What techniques should be used to measure the properties selected for evaluation?

Statement of the Objectives

This research effort has two objectives:

1. Determine the criterion variables most applicable to an initial flight simulator evaluation of effectiveness.

2. Develop a general technique for the initial evaluation of flight simulator effectiveness.

Research Methodology

A literature search of current military flight simulator evaluation techniques was conducted. This search was limited primarily to those techniques documented in the Defense Documentation Center. The using agencies examined included the Army, Navy, and Air Force. In addition, current aircraft test and evaluation techniques were reviewed for possible application to flight simulator evaluation. Applicable Air Force regulations were also studies in order to obtain a more complete understanding of the current Air Force test and evaluation process.

The most suitable properties for evaluation had to be determined prior to the development of an initial evaluation technique. The first step in this determination was to define those flight simulator properties which could possibly be used as evaluation criterion variables. Second, each possible property then had to be examined to determine if the relationship between simulator effectiveness and the property to be measured was well defined and meaningful. In addition, for a simulator property to qualify as a criterion variable, techniques had to be available to accurately measure the simulator property within the environment of an initial flight simulator evaluation. The information required to select simulator properties for evaluation was extracted from the current literature and through informal discussions with the personnel of the Simulator System Program Office at Wright-Patterson Air Force Base, Ohio.

Once the most suitable properties were determined, the methodology transitioned to development of the initial flight simulator evaluation technique. This part of the methodology involved the combination of two research methods. The first was the literature search of existing flight simulator evaluation techniques and the second was personal observations of an initial flight simulator evaluation which was on-going during the research effort.

At the time this research effort began, an initial evaluation of the T-37 Undergraduate Pilot Training Instrument Flight Simulator (UPT-IFS) was in the planning stages. The planned evaluation included a more traditional quantitative evaluation plus the use of qualified T-37 instructor pilots to obtain a subjective evaluation of the potential effectiveness of the T-37 UPT-IFS. A questionnaire had been constructed by Air Training Command (ATC) for use in the collection of data during the subjective portion of the evaluation. The questionnaire and data collection methods were examined. Modifications were recommended during this research effort in an attempt to improve the evaluation. Observations during a portion of this evaluation were analyzed to determine the potential effectiveness of the subjective technique. The analysis was directed toward the identification of modifications that would improve the usefulness of the combined quantitative and subjective evaluation techniques.

The information acquired from the T-37 UPT-IFS was then combined with the favorable characteristics of the current techniques examined during the literature search. The resulting compilation was a generalized initial evaluation technique for estimating flight simulator effectiveness. Hopefully, this technique will be useful during the initial evaluation of flight simulators in future Air Force procurement actions.

Research Scope and Limitations

The scope of this research effort was limited to a small portion of the overall evaluation of a flight simulator. The complete evaluation includes estimates of military utility, operational effectiveness, compatibility, interoperability, reliability, maintainability, logistic supportability, cost of ownership, and training requirements (AFR 80-14, 1975). This research effort focused on the operational effectiveness, or performance, of the flight simulator.

Many authorities in the flight simulation field support the hypothesis that the design of the training program in which a flight simulator is used is at least as important to simulator effectiveness as is the design and performance of the simulator itself (Caro, 1973). Although this hypothesis appeared to be valid, it was of little assistance for an initial simulator evaluation. Good training programs are the result of a substantial amount of experience using the equipment. This research did not address simulator training programs but instead, concentrated on the performance of the flight simulator equipment as a separate entity.

An initial flight simulator evaluation is restricted by the environment in which it must be performed. The initial evaluation usually is conducted in the flight simulator contractor facility. Usually, the instructors, operators, and students who will eventually use the equipment are not available in the immediate area of the evaluation. Budgetary limitations on temporary duty funds normally restrict the number of military personnel available to participate in the evaluation.

The flight simulator to be tested is normally set up on a temporary site which is usually space limited. Frequently, cost considerations require that some intended features of the eventual flight simulator be omitted in construction of the test device. For example, only two of the four cockpits made available for the initial T-37 UPT-IFS evaluation were equipped with visual displays. The simulators delivered to ATC will have visual displays on all four cockpits. Thus, cost considerations as well as space limitations frequently lead to degrading of the test device.

Simulator development and production are almost always on a rigid schedule. Frequently, problems develop that cause the contractor to get well behind the schedule. This slippage often results in attempts to compress the final events of the schedule, in which test and evaluation are usually found.

<u>Summary</u>. This research effort treated only one of the many factors involved in a flight simulator evaluation. That factor was simulator effectiveness or performance. The research was also limited to an initial flight simulator evaluation and therefore did not deal with the design of the training program used with the simulator. The environment of the initial evaluation also placed several limitations on the research effort and on the resulting initial flight simulator evaluation technique.

Thesis Organization

The remaining chapters of this report develop the criterion variables for evaluation and an initial evaluation technique. Chapter II examines the techniques currently in use for military flight simulator evaluations. Chapter III is an examination of the properties of flight simulators which should be considered as evaluation criterion variables. Chapter IV develops an evaluation technique for the initial estimation of flight simulator effectiveness. Chapter V presents the conclusions and summarizes the general technique for the initial evaluation of a flight simulator.

II. Current Evaluation Techniques

Introduction

In order to avoid reinvention of the wheel, this research effort begins with an examination of simulator evaluation techniques that have been or are being used on modern flight simulators. The examination has been limited to military evaluation techniques because of the unique requirements of military flying and military flight simulation. Current evaluation techniques used by the Navy, Army, and Air Force are examined. The examination is based primarily on evaluations conducted since 1970 that have been documented in the Defense Documentation Center. This examination identifies many good techniques to build on and several poor techniques to be avoided.

Navy Flight Simulator Evaluation Techniques

The evaluation techniques used during two recent procurement actions are representative of the techniques currently used by the Navy. Navy Device 2F101 was evaluated from 16 August 1973 to 24 October 1974 during initial acquisition (Walker & Galloway, 1975). Device 2F90 was evaluated from 18 March 1974 to 4 December 1974 following the addition of a computer generated visual display system (Galloway & Hewett, 1975). This device had to be evaluated again during 9 to 13 June 1975 (Hewett & Galloway, 1975). Device 2F101 simulates the T-2C aircraft and Device 2F90 simulates the TA-4J aircraft (Hewett, Galloway, & Murray, 1974).

Device 2F101. The purpose of the evaluation of Device 2F101 was to assist the contractor (Singer Simulation Products Division of Binghamton, New York) with the development of a T-2C flight simulator. The initial design of flying characteristics in Device 2F101 was based on wind tunnel estimates of the various T-2C aircraft characteristics. These estimates focused on the edge of the aircraft flight envelope and proved to be inadequate for programming the T-2C flight characteristics (Walker & Galloway, 1975).

The approach taken to improve the simulation of flying characteristics in Device 2F101 was to generate a more complete and accurate data base for the T-2C aircraft characteristics and then use this improved data base for the design of a mathematical model which could be programmed in the flight simulator computers. The Naval Air Test Center (NATC) team began this approach by flying 44 sorties in 7 different T-2C aircraft for a total of 63.3 flight hours (Walker & Galloway, 1975).

Each of the variables which describe the flying characteristics of the T-2C aircraft was recorded during all normally encountered combinations of configuration and environmental conditions in each of the seven aircraft. The instrumentation used to measure these variables consisted of: production instrumentation, a sensitive airspeed indicator, calibrated altimeter, calibrated angle of attack system, sensitive accelerometer, a 0-to 100-pound hand held force guage, a 3-second sweep stopwatch, a 48-inch tape measure, and a sensitive inclinometer.

The values of each variable were then averaged over the seven aircraft flown to produce an estimate for the average T-2C aircraft. The resulting data base mapped the T-2C flight characteristics over the entire operational flight envelope of the aircraft (Walker & Galloway, 1975). Rather than develop a new mathematical model for programming the flight simulator computers, the NATC team elected to make similar measurements on Device 2F101 and conduct a comparison of the two data sets.

The application of corrections became an iterative process within three separate phases. The separate phases in which the corrections were made involved: (1) the fixed base flight simulator with motion and visual systems disconnected; (2) the fixed base simulator with only motion connected; and (3) the total flight simulator with both motion and visual systems connected. The purpose of this three-phase approach was to simplify the corrections by reducing the number of system interactions involved at any one point in the evaluation.

The need for an iterative process within each phase was also generated by the interaction between variables. All flight characteristics had to be checked after each correction to insure that the new correction did not destroy the effects of previous corrections. If this occurred, both corrections had to be compromised in order to minimize the impact of the remaining uncorrectable deficiencies.

The NATC team also found that the test pilot had to fly the actual aircraft frequently during the simulator evaluation in order to maintain his objectivity. The test pilot quickly adapted to the simulator characteristics and needed to fly the aircraft to reconfirm the differences between aircraft and simulator characteristics. As an additional precaution against this tendency, fleet pilots currently qualified in the T-2C aircraft were requested to fly and evaluate the simulator on several occasions during the evaluation (Hewett, Galloway, & Murray, 1974).

Device 2F90. The evaluation of this flight simulator began seven months after the evaluation of Device 2F101. Device 2F90 had been in use since 1969. It was originally purchased with a motion system but without a visual display system. The fidelity, or accuracy, of simulation was judged by the users to be barely adequate enough to make the simulator usable for training (Galloway & Hewett, 1975). The addition of a computer generated visual display accentuated the shortcomings of the device to the extent that corrective action became necessary (Harris, 1975).

The Naval Air Test Center team used an approach which was almost identical to the one used for Device 2F101. The first problem addressed was the lack of a sufficient data base to describe the flying characteristics of the TA-4J aircraft. The data collection effort for this evaluation was on a much smaller scale than the one conducted for Device 2F101. Only 20.5 hours were flown in the eleven aircraft measured.

The accuracy and cost of the instrumentation used for measurements in the aircraft was greatly reduced. Expensive production instrumentation was not used. The instrumentation used for aircraft measurements consisted of: a hand held force gauge, a tape measure, a stopwatch, and the uncalibrated cockpit instrumentation common to all TA-4J aircraft (Galloway & Hewett, 1975). The NATC test team obviously felt that the reduction in the accuracy of data collected was insignificant when compared to the increased cost of fitting the eleven aircraft with production instrumentation.

Each required variable was measured in each of the eleven TA-4J aircraft. These measurements were then averaged as in the Device 2F101 evaluation. In an attempt to reduce the bias in the data, all measurements were taken by the same pilot using consistent measuring techniques.

The variables measured in the aircraft were also measured in the flight simulator. The simulator measurements were taken with the same configuration, environmental conditions, pilot, and measurement techniques as were present during the corresponding aircraft measurements of each variable. The aircraft and flight simulator data bases were then compared.

The comparison and correction planning were performed in a different manner than for Device 2F101 since the flight simulator had been in operational use before the evaluation. The prior years of use had identified the specific problem areas in the characteristics of the flight simulator. Because of this, it was possible to select discrepancies between the two data sets which were probable causes of the identified problems. These selected discrepancies were the only ones considered during planning of the corrective actions (Harris, 1975).

The corrective action for Device 2F90 followed the same pattern as was used for Device 2F101. However, the pattern of corrective action used for Device 2F90 consisted of a purely quantitative comparison of the two data sets. After all corrective action was completed, a very informal subjective evaluation of Device 2F101 was performed by several instructor pilots who were currently qualified in the TA-4J aircraft. Very little emphasis was placed on this portion of the evaluation. This is indicated by the following two-sentence summary contained in the NATC report:

> The instructor pilots evaluating the simulator were impressed by the fidelity of the total simulation and agreed that training substitution could be realized in VFR (Visual Flight Rules) flight simulation with Device 2F90. Some of the evaluating pilots experienced lateral PIO (Pilot Induced Oscillation) tendencies and thought that laterally the simulator was too sensitive. (Galloway & Hewett, 1975, p 32)

These findings were quickly dismissed by the decision of the test team to delay any additional fidelity improvements until operational use confirmed that deficiencies still existed (Galloway & Hewett, 1975).

It only took eight months (October 1974 to June 1975) of operational use to confirm the findings of the informal subjective evaluation. At that time, the Chief, Naval Air Training Command, requested that the test team return in order to correct the deficiencies identified during operational use. The test team began the new evaluation by analyzing the comments which had been written by instructor and student pilots. This analysis was used to design an appropriate test technique for deficiency correction (Hewett & Galloway, 1975).

In addition to the analysis of written comments, approximately 40 cockpit hours were flown to evaluate corrections as they were made. All of the corrections made during this evaluation involved the addition of variables that had been omitted from the previous mathematical models and flight simulator computer programs. The test team concluded that this second evaluation had minimized the deficiencies identified by the users within the limitations of the system hardware and software (Hewett & Galloway, 1975). No additional reports are available to either verify or contradict this conclusion.

Summary of Current Navy Evaluation Techniques. The general technique currently used by the Navy to evaluate and improve a flight simulator follows five sequential steps. These are: (1) gather base line data from the aircraft; (2) perform the same flight tests in the simulator; (3) reduce and compare the two data sets and plan the corrective actions to be made; (4) apply the corrections; and (5) repeat as necessary.

The base line data from the aircraft are measured in several different aircraft and then averaged. Simple and inexpensive instruments can be used for these measurements. A hand held force gauge, stopwatch, tape measure, and standard cockpit instruments are sufficient.

The initial measurement of variables describing flight characteristics in the simulator is made under the same environmental conditions that existed during the aircraft measurements. In addition, these measurements are taken with both the motion and visual systems disconnected if the simulator is so equipped.

The data analysis should be accomplished by comparisons of the aircraft and flight simulator data sets. If specific simulator problems have previously been identified, the comparison concentrates on the discrepancies which are probable causes of these problems. Corrections are planned in advance to minimize the impact of variable interactions.

The corrections are made in three separate phases: (1) on the fixed base simulator alone; (2) with only the motion system connected; and (3) with both the motion and visual systems connected. Within each phase, an iterative process is used. After each correction is made, all variables are rechecked to insure that the impact of variable interactions has been correctly assessed. The test pilot is allowed to fly the aircraft frequently and currently qualified fleet pilots are used to prevent inaccurate measurements caused by rapid orientation to the simulator characteristics.

Army Flight Simulator Evaluation Techniques

Two recent flight simulator evaluations conducted by the Army are examined in this section. The first evaluation was on Device 2B24 which simulates flight in the UH-1H helicopter. The approach used for this evaluation has been described as the traditional Army approach to flight simulator development and evaluation. The second evaluation was conducted for Device 2B31 which simulates flight in the CH-47 helicopter. The approach used here was described as the new Army approach (Catron, 1975). The evaluation of Device 2B24 was conducted during 1971 and 1972 (Caro, Isley, & Jolly, 1975). The exact time of the evaluation on Device 2B31 was not specified in the report. However, references to the evaluation on Device 2B24 indicate that this evaluation was more recent (Catron, 1975).

Device 2B24. The development of the UH-1H flight simulator represents the traditional Army approach to development and evaluation of a flight simulator. The contractor was responsible for acquisition of the aerodynamic data used for the simulation. The only data source available to the contractor was the coefficient data obtained during the UH-1A helicopter development. These data were based on engineering predictions and wind tunnel testing of the UH-1A components. New coefficient data were not generated for the five model changes between the UH-1A and the UH-1H even though some major aerodynamic changes were made (Catron, 1975). Therefore, the data used for development of the UH-1H helicopter simulator were actually estimates based on the characteristics of the UH-1A helicopter.

The contractor used the coefficient data to develop an off-line computer program for computation of the performance and flying qualities of the UH-1H helicopter. This program was an exact model of the UH-1A data collected. The accuracy of this model had to be compromised when it was converted into a real time program which could be used in the flight simulator computers. This reduction in accuracy is due primarily to software limitations (Catron, 1975). The relationship between the data sets discussed here is symbolized in Figure 2-1.

It should be noted that the data used for programming the UH-1H simulator has a questionable relationship to the actual UH-1H flying characteristics. It is actually only an approximation of the estimated

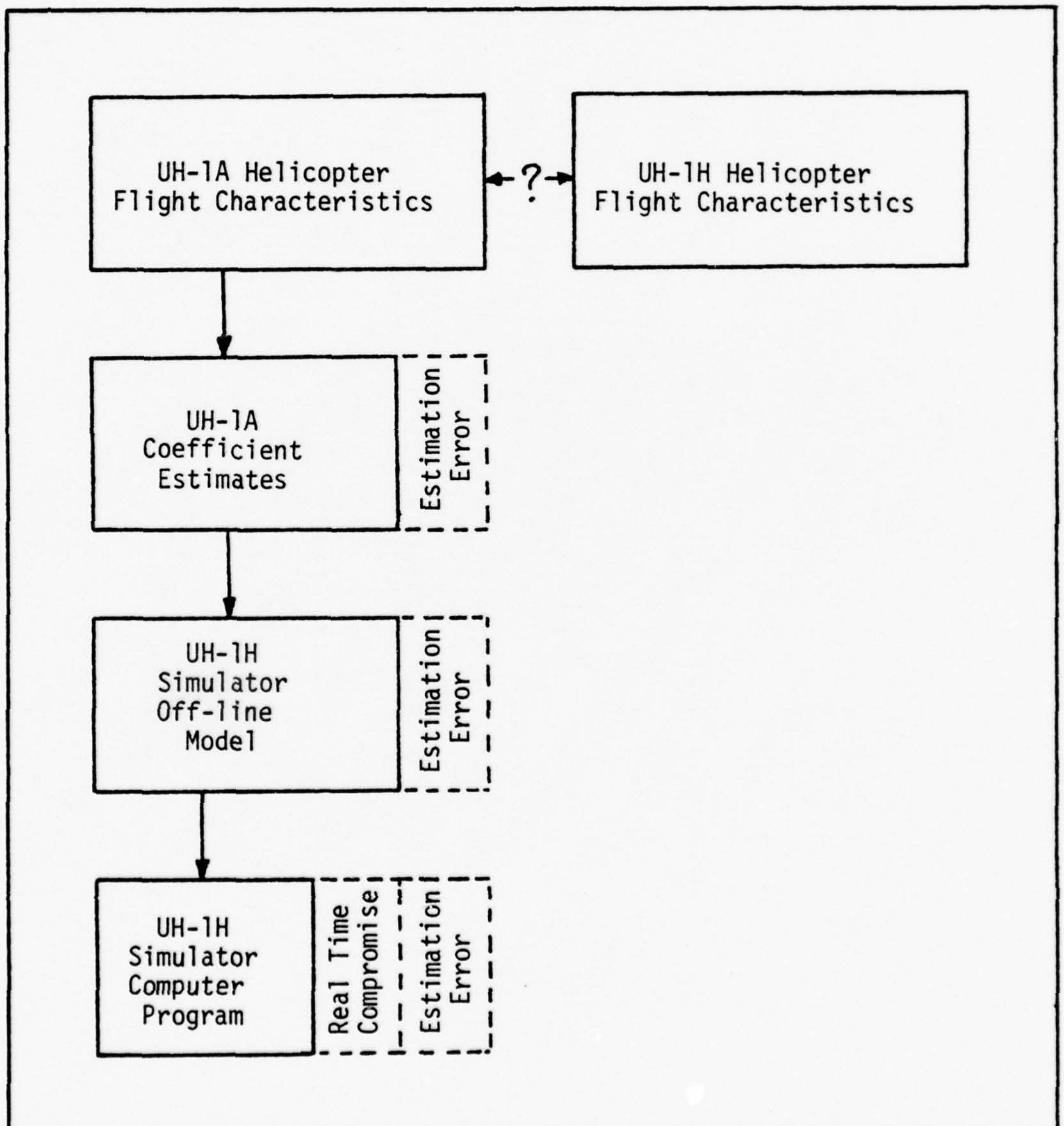


Figure 2-1. Derivation of the UH-1H Simulator Model

characteristics of the original UH-1 helicopter. This point is stressed here since it seems to be a common problem to all flight simulator programs.

The initial evaluation of Device 2B24 was purely quantitative in nature. Tolerances were developed around the UH-1H simulator off-line model described above. The simulator coefficients were then compared with the off-line model. When all simulator coefficients were within the developed tolerances, the simulator passed the evaluation (Catron, 1975). This type of evaluation only insures that the UH-1H simulator is a good approximation the model which represents an estimate of the UH-1A helicopter flight characteristics. The qualities of this traditional Army evaluation are best discussed by examining a second evaluation which was conducted after Device 2B24 passed this initial evaluation and was delivered to the Army.

The Army Aviation Test Board was responsible for conducting Expanded Service Tests after delivery of the flight simulator. The board requested the Human Resources Research Organization (HumRRO) develop an evaluation plan that would determine the mission suitability of the device (Caro, Isley, & Jolly, 1975).

HumRRO developed a three-phase evaluation for this purpose. Phase I activities were devoted to evaluating the workability of the device and assessing the fidelity or accuracy of simulation of the UH-1H helicopter. Phase II designed a training program which made optimum use of the training features included in the device. Phase III efforts were directed toward estimating the transfer of training and cost effectiveness of Device 2B24 (Caro, Isley, & Jolly, 1975). Phases II and III of the evaluation plan are not applicable to the subject of this research effort. Phase I, the evaluation of performance, will be examined for application to the development of an initial evaluation of flight simulator effectiveness and for assessment of the traditional Army approach to simulator evaluation.

The plan for evaluation of performance developed by HumRRO was divided into three parts. The first portion of the evaluation was devoted to test staff familiarization with the equipment and its operation. This portion of the plan was intended to quality the test staff in the use of the controls unique to Device 2B24 before the evaluation was conducted. Following the familiarization, the plan required mock training be conducted by the test staff in several cockpits simultaneously. Performance assessment was to be made by the staff during this portion of the evaluation. The final portion of the plan was designed to use a relatively large number of Army aviators from the local area for evaluation. It was intended that these aviators fly the device and rate the fidelity or accuracy of simulation and acceptability of the device for training. The pilots were to rate the simulator by responding to a questionnaire and structured interview (Caro, Isley, & Jolly, 1975).

Unfortunately, the evaluation plan developed by HumRRO could not be followed during Phase I. During the initial familiarization activities, it became obvious that the device still had major deficiencies that prohibited the mock training and subjective pilot ratings contained in the evaluation plan. Phase I activities became an effort to identify deficiencies and effect corrections so that the device could become suitable for training. Because of the many deficiency corrections being made and the limited availability of the device during those corrections, the test activities had to be substantially reduced. It was not practical to schedule any nontest staff personnel to fly the device. Therefore, all data on the accuracy of simulation and device acceptability were obtained from test staff members (Caro, Isley, & Jolly, 1975).

Because of these circumstances, all data collected during the performance evaluation were informal opinion data. The data collection method adopted required two test staff members who were UH-1H qualified pilots to fly the simulator together. When a discrepancy was detected, a judgement was made jointly by the pilots and the HumRRO staff as to the probable impact on the training capabilities of the simulator. The questionnaire and structured interview developed during the evaluation planning were not used (Carol Isley, & Jolly, 1975). In effect, the Expanded Service Test had to do the tasks that the initial evaluation should have accomplished.

Device 2B31. The new approach used for development and evaluation of the CH-47 helicopter simulator follows part of the traditional approach used on Device 2B24. The contractor was again responsible for collection of the flight characteristics data. These data were again estimates of the flight characteristics which were obtained from the helicopter manufacturer. The simulator was designed and programmed based on this estimated data base. After initial development, the new approach departed from the traditional approach in three important areas.

First, the Army designated a team of two pilots to provide preliminary evaluations prior to the beginning of actual acceptance testing. The pilot designated as team leader participated in all three of these preliminary evaluations. A different pilot assisted the team leader in each preliminary evaluation. The pilot evaluation team flew the simulator and wrote Discrepancy Reports (DR's) on all simulator characteristics which were perceived to be different from the actual CH-47 helicopter. The original data base was changed after each preliminary evaluation to eliminate or minimize the identified deficiency. The key to the success of this approach was that evaluation team comments and changes took precedence over the original data base collected from the manufacturer (Catron, 1975).

The second departure from the traditional approach involved flight time in the CH-47 helicopter for the contractor personnel. The purpose of these orientation flights was to allow contractor personnel to gain first-hand experience with the helicopter flight characteristics and the pilot techniques used to fly the helicopter. The evaluation pilots were of the opinion that this first-hand experience would improve the communications and understanding between the contractor and the Army (Catron, 1975).

The third important change in the new approach involved the standards used for acceptance testing. The traditional approach compared the simulator characteristics with the off-line model developed from the original estimates of the helicopter characteristics. The new approach used the model that resulted from the extensive modifications made during the preliminary evaluations of the first device as the standard for all follow-on devices. This change meant that the simulator performance was compared to the actual CH-47 helicopter performance as it was perceived by the evaluation pilots (Catron, 1975).

Summary of Current Army Evaluation Techniques. The new approach used for development and evaluation of Device 2B31 recognized that the real world aircraft data base is normally inadequate for programming a flight simulator. The approach developed made use of qualified CH-47 helicopter pilots to correct the data base deficiencies. The key to the success of this approach was the recognition that evaluation pilot comments had to take precedence over all other sources of coefficient data.

The Army concept of a two-pilot evaluation team solved two problems characteristic of a subjective evaluation. The team leader added consistency to the evaluations since he participated in all evaluations conducted. The use of a different assistant pilot for each evaluation made new inputs available and helped in the identification of deficiencies which the team leader had adapted to and no longer recognized. This same problem of a pilot adapting to the simulator characteristics was identified in the Navy evaluations reviewed.

Air Force Flight Simulator Evaluation Techniques

Major revisions have been made in the Air Force flight simulator acquisition process since 1970. The most significant change was the centralization of control for flight simulator acquisition in the organization of the Simulator System Program Office at Wright-Patterson Air Force Base, Ohio in 1973. This organization has served as the focal point for the rapidly increasing emphasis on the development and use of flight simulation in the Air Force (Dunlap, et al., 1975).

One of the acquisitions currently being managed by the Simulator System Program Office is the Undergraduate Pilot Training Instrument Flight Simulator (UPT-IFS). A new Air Force approach is being used for development and evaluation of UPT-IFS. This new approach is discussed in Chapter IV which is devoted to the subjective evaluation of flight simulator effectiveness. The following discussion will concentrate on the more traditional Air Force approach that has been used prior to UPT-IFS.

Although prior Air Force flight simulator evaluations have not been documented in the Defense Documentation Center, two very general articles discuss the technique used and some of the problems encountered. The technique discussed in these articles has been labeled "tweaking" (Richmond, 1976).

The process of "tweaking" a flight simulator is almost identical to the new approach used by the Army. The simulator is designed and constructed with the use of estimated data for the aircraft flight characteristics. After construction and programming are complete, qualified pilots fly the simulator and attempt to identify simulator characteristics which are perceived to be different from the real world aircraft. The simulator contractor then changes the programmed data base until the simulator characteristics in question are perceived by the pilots to be the same as the aircraft characteristics (Rust, 1975).

Several problems have been encountered while using this "tweaking" technique. First, the pilot does not always perceive the real problem. For example, the pilot may perceive that the stick foreces for the roll axis are too heavy for a given airspeed. In actuality, this perception could be caused by other factors such as incorrect stick deflections for a given force. If the contractor simply changes the stick forces, the problem will probably still exist. In other words, a pilot is capable of identifying a deficient area but does not usually have sufficient engineering and programming expertise to recommend a specific corrective action for the problem (Richmond, 1976).

The second common problem with this approach is that while the pilot is "tweaking" the simulator, the simulator is "tweaking" the pilot. For example, in the stick forces problem discussed earlier, suppose the forces were actually five pounds too heavy. Now, suppose the contractor reduces these forces to only two pounds too heavy. It is very probable that the pilot will perceive the new forces as correct after having experienced the heavier forces prior to the correction. If the pilot flys the real world aircraft and then returns to the simulator, he will probably perceive the forces as being too heavy again. The contractor will surely begin to doubt the pilot's ability to accurately perceive control stick forces. In effect, the simulator has "tweaked" the pilot (Rust, 1975).

The last major problem with "tweaking" is a result of the differences between the perceptions of individual pilots. If the contractor manages to satisfy one pilot with the simulator characteristics, a different pilot may require the iterative process to start from the beginning in order to produce characteristics that agree with his perceptions. As a result, the "tweaking" technique is usually not repeatable for different pilots (Richmond, 1976).

The Air Force recognizes that "tweaking" is a slow, iterative, and frequently nonrepeatable process. A more ordered Air Force approach to simulator evaluation is being developed for UPT-IFS. One of the objectives of this research effort is to aid in the formulation of this new approach which is discussed in greater detail in the remaining chapters of this report.

Summary of Current Evaluation Techniques

The main problem identified by each of the military agencies was a lack of sufficient data to completely and accurately describe the flight characteristics of the aircraft being simulated. Normally, the only data available were the manufacturer's estimates for the development model of the aircraft series. Additional accuracy is lost because of software limitations which prohibit programming an exact model of the data that is available.

Generally, two approaches have been taken to reduce the impact of this problem. The Navy attempts to obtain the required data for effective simulation by extensive flight testing of the real world aircraft. The Army approach and the previously used Air Force approach attempt to compensate for data deficiencies through pilot perceptions of the simulator and aircraft characteristics.

The purely quantitative approach used by the Navy does not appear to be adequate in the identification of deficiencies that will impact on the eventual effectiveness of the flight simulator. The subjective techniques used by the Army and the Air Force tend to be very slow, iterative, and nonrepeatable. However, the Army concept of an evaluation team with a team leader seems to improve the consistency of this technique.

Almost all of the techniques examined here were aimed at an evaluation of fidelity or accuracy of simulation. The validity of using fidelity and several other criterion variables for flight simulator evaluation is discussed in Chapter III.

III. Criterion Variables for Flight Simulator Evaluation

Introduction

An evaluation is an examination that results in a judgement about the worth or quality of the item being evaluated (Webster, 1963). This general definition implies that before an evaluation can be conducted, the evaluator must know: (1) the kind of quality desired in the item; (2) the factors that the judgement of quality will be based on; and (3) the methods that will be used to examine the factors. These three aspects of an evaluation are not independent of each other. The kind of quality desired will dictate which factors will be used for a basis of judgement. These factors will then determine the method of examination that can be used.

This chapter discusses the first two aspects of an evaluation for Air Force flight simulators: (1) the kind of quality desired in a flight simulator, and (2) the factors that the judgement of quality will be based on. The third aspect, the method to be used for examination, is discussed in Chapter IV.

Chapter Organization

The first two sections of this chapter are devoted to an analysis of the kind of quality desired in an Air Force flight simulator. The approach taken for this analysis is to first review the objectives of flight simulator test and evaluation. Since these objectives are stated in terms of the mission of Air Force flight simulators, the second section examines this mission. These two sections, taken together, provide the background necessary for the selection of the criterion variables to be measured during the evaluation.

The remainder of the chapter is devoted to an examination of the criterion variables most frequently measured during a flight simulator evaluation. This examination results in a conceptualization of the relationships among possible criterion variables and in the selection of a criterion variable which provides the best information available for the judgement of quality required in an initial flight simulator evaluation.

Flight Simulator Evaluation Objectives

During the acquisition of a new Air Force flight simulator, there are two types of test and evaluation performed. Development Test and Evaluation is conducted to assess the accomplishments of the development phase and to provide guidance for the remaining development effort. Operational Test and Evaluation is usually performed by the using command to estimate the effectiveness the device will have after delivery and then to actually measure the effectiveness after the device is in operational use.

Operational Test and Evaluation (OT&E) requires several years to complete. The entire OT&E process can be divided into Initial Operational Test and Evaluation (IOT&E), which is performed prior to delivery of the flight simulator, and Follow-on Operational Test and Evaluation, which is performed during operational use after delivery of the device to the using activity (AFR 80-14, 1975).

The objective of this research effort is to develop an evaluation technique which will provide a valid estimate of simulator effectiveness prior to delivery of the device. This function is currently being performed during Initial Operational Test and Evaluation. Therefore, the objectives of this evaluation will be examined more closely.

An Initial Operational Test and Evaluation, as defined by Air Force Regulation 80-14, is:

> That test and evaluation performed during a development program intended for acquisition. It is an initial phase of operational test and evaluation adequate to provide, prior to the first major production decision, a valid estimate of expected system operational effectiveness and suitability. (AFR 80-14, 1975, p 13.)

Therefore, the objective for an IOT&E is to provide a valid estimate of the expected operational effectiveness and suitability.

Operational suitability consists of many areas such as compatibility, interoperability, maintainability, logistics supportability, cost of ownership, and training requirements (AFR 80-14, 1975). Good estimating techniques are already available for these suitability areas and will not be included in the evaluation developed by this effort. Instead, the evaluation developed here will concentrate only on an estimate of the operational effectiveness of the device.

Operational effectiveness is how well the system performs its intended mission when operated in its intended environment (AFR 80-14, 1975). Criterion variables for the evaluation must provide enough information to judge how well the simulator will perform its intended mission. Therefore, the mission of Air Force flight simulation must be examined to determine which variables are to be selected for estimating the capability of a simulator to perform this mission.

Intended Mission of Air Force Flight Simulators

There are two distinct uses for flight simulators in the Air Force today. They are used for: (1) the evaluation of engineering and human factors, and (2) aircrew training (see Figure 3-1). Flight simulators used for the evaluation of engineering and human factors are usually developed and operated by elements of Air Force Systems Command (AFSC) for investigating of such matters as flight instrument display and layout, aircraft design performance comparisons, stability and control criteria, and so on (Dunlap, et al., 1975). Flight simulators used for aircrew training are generally developed by AFSC, but are used by other major commands as operational trainers. This research effort is limited to those flight simulators intended for use as aircrew training devices. Throughout the remainder of this report, the term "flight simulator" will mean only flight simulators intended for use in aircrew training.

As discussed in Chapter I, the mission of Air Force flight simulators has changed frequently during the past few years. Changes in the

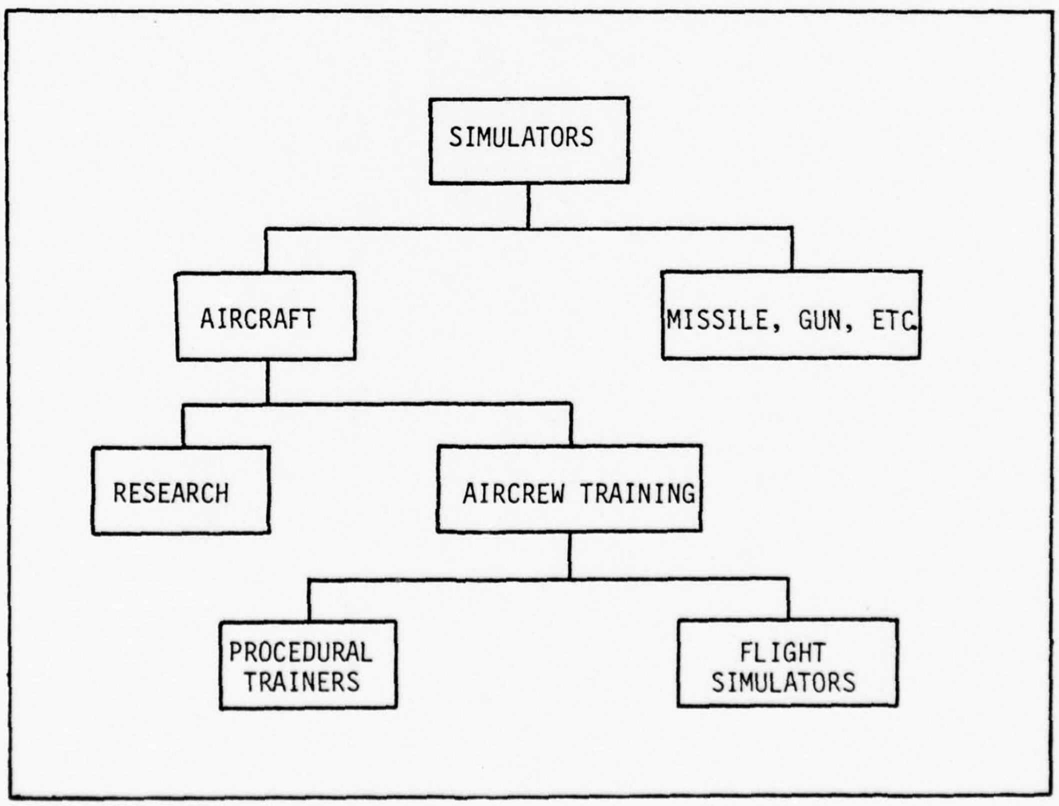


Figure 3-1. Categories of Air Force Simulators

intended mission have been the result of technological advances in aircraft or simulator design as well as external forces acting on the aircrew training environment.

The most significant technological advance in simulator design was the introduction of the digital computer. The high capacity and accuracy of these machines greatly improved simulation of aircraft handling characteristics and made it possible to add visual and motion systems. These improvements extended the capability of flight simulators to many Visual Flight Rule (VFR) areas that had traditionally been limited to training in the aircraft.

The external political and economic forces discussed in Chapter I required these extended capabilities to be included in Air Force flight simulators. The Chief of Staff of the Air Force stated this requirement in his message of 25 April 1975. The message, in part, was:

> Air Force policy is to strive for a 25 percent reduction in flying hours by the end of FY 81 through the increased use of simulation. While operating costs and energy considerations are the driving factors, other reasons such as restricted airspace, environmental ecological impacts, safety and aircraft attrition are also major considerations. (Dunlap, et al., 1975, p 136)

The required reduction in flying hours implies that the use of flight simulators for teaching and maintaining flying skills used in instrument conditions must be expanded. In addition, the flight simulators must be capable of assuming some of the training for flight in VFR conditions which has traditionally been conducted in the aircraft.

Therefore, the current mission of Air Force flight simulation is to reduce aircraft flying time used for training by providing an effective environment in which flying skills can be taught and maintained for both instrument and VFR conditions.

Possible Criterion Variables

The discussions of the two previous sections imply that an initial flight simulator evaluation should provide a valid estimate of how well instrument and VFR flying skills can be taught and maintained in the environment provided by the flight simulator. The quality of training in a flight simulator is a composite of the efficiency of original learning, the transfer of what was learned in the simulator to performance in the aircraft, and the retention of what was learned (Williges, *et al.*, 1972). Each of these elements of training are dependent on many variables, only a few of which are related to the characteristics of the flight simulator. Very little is known about which variables have the most important impact on the quality of training. Therefore, the approach which must be taken is to select the criterion variables which will give the best estimate of the quality of the flight simulator.

The remainder of this chapter is devoted to an examination of the individual criterion variables available for measurement during an initial flight simulator evaluation. Before discussing each criterion variable separately, the terminology to be used for the discussion and the relationship between criterion variables must be defined. The key definitions for this section are:

Training Efficiency - A measurement of the efficiency of original learning which takes place within the flight simulator (*Training Devices and Simulation: Some Research Issues*, 1954).

Transfer of Training - A measurement of how training conducted in the flight simulator influences subsequent performance in the aircraft represented (Ellis, 1965).

Fidelity of Engineering Simulation - A measurement of how well the physical characteristics of the real world aircraft have been copied in the flight simulator (Miller, 1954).

Fidelity of Psychological Simulation - A measurement of how the thought processes generated by training in the flight simulator affect the thought processes required for performance in the aircraft (Miller, 1954). Fidelity of psychological simulation includes the concepts of training efficiency and transfer of training. It is also equivalent to flight simulator effectiveness.

Simulator Effectiveness - A measure of the quality of the flight simulator. Simulator effectiveness describes how well the flight simulator will perform its intended mission of reducing aircraft flying time used for training by providing an effective environment in which to teach and maintain flying skills in instrument and VFR conditions.

The relationships between these terms can be summarized by the following set of equations:

Simulator Effectiveness ≡ Aircraft Flight Time Saved + Quality of Training
≡ Training Efficiency + Transfer of Training
≡ Fidelity of Psychological Simulation
≡ f (Fidelity of Engineering Simulation)

In words, simulator effectiveness can be measured in units of aircraft flight time saved, through use of the flight simulator, plus the quality of training provided by the combination of training in the actual aircraft and the flight simulator. It can also be measured by the training efficiency of the original learning in the flight simulator plus the amount of that training which transfers to performance in the aircraft.

Simulator Effectiveness is considered to be equivalent to Fidelity of Psychological Simulation and is a function of the Fidelity of Engineering Simulation.

The next sections discuss each of the possible criterion variables and rationalize the above equivalences between the possible criterion variables.

Aircraft Flight Time Reductions

The mission of Air Force flight simulators is to reduce aircraft flying time by providing an environment in which flying skills can be effectively taught and maintained. This mission implies that simulator effectiveness could be partially measured by units of aircraft flight time saved. In order for the units of aircraft flight time saved to be meaningful, the resulting quality of training must be the same as it was before the amount of aircraft flight time was reduced.

It was mentioned earlier that quality of training is too complex an issue to be measured directly because of limited knowledge about the variables which affect it. Since the quality of training cannot be measured directly, holding the quality of training constant during a flying time reduction is all but impossible.

A measurement of simulator effectiveness in terms of flying time saved is frequently attempted during Follow-on Operational Test and Evaluation. The normal approach is to train a control group of student pilots without the flight simulator and an experimental group with the simulator. The two groups are trained to the same performance level and the difference in aircraft time used represents the simulator effectiveness (Caro, Isley, & Jolly, 1975). This approach is limited by problems common to all controlled experiments. The initial ability of the student pilots must be evenly distributed between the two groups. The difference in training quality due to different training programs and different instructor pilots must be isolated and a means must be available to accurately assess the performance level of each student pilot at the end of the experiment (Williges, et al., 1972).

Much research has been devoted to solving these many problems associated with the performance rating approach for direct measurement of simulator effectiveness. So far, little progress had been made (Williges, et al., 1972). When the limitations of an initial flight simulator environment are imposed on this approach, it becomes impossible to measure simulator effectiveness in terms of aircraft flying time saved. However, once simulator effectiveness has been estimated, an approximation of flying time saved can probably be extracted from the estimate.

Training Efficiency

Training efficiency is a measure of the efficiency of the original learning that takes place during training in the simulator. It is measured by changes in performance for a given amount of simulator training conducted (Training Devices and Simulators: Some Research Issues, 1954). For example, suppose that a 60-degree bank turn is being taught to two students of equal ability in two different flight simulators. If it takes three hours in the first simulator and four hours in the second simulator before the students reach a satisfactory level of performance on 60-degree bank turns, it could be concluded that the first simulator has more training efficiency that the second.

Unfortunately, a measurement of time to achieve a given performance level is not possible within the environmental limitations of an initial flight simulator evaluation. Even if it were possible, the problems of the performance measurement technique discussed in the previous section would greatly reduce the accuracy of the approach.

This researcher supports the hypothesis that it is possible to estimate the training efficiency of a flight simulator through the use of a subjective evaluation by instructor pilots currently qualified in the aircraft being simulated. This hypothesis is based on the assumption that the accumulated experience of instructor pilots qualifies them to judge how well a student pilot will learn in a given simulated environment. This expertise should be capable of providing an accurate estimate of the training efficiency a flight simulator will have in operational use. A subjective evaluation of this type is possible within the limited environment of an initial flight simulator evaluation and will be discussed in greater detail in Chapter IV.

Transfer of Training

As mentioned in the previous section, training efficiency is only a measure of the efficiency of original learning. This measurement is only one part of the overall simulator quality. From the example of the 60-degree bank turns, it was concluded that the simulator in which only three hours of training were required had better training efficiency. However, this superior training efficiency is meaningless if the skills required to perform this maneuver in the simulator are completely different from those required in the aircraft simulated. In this case, the simulator training would not improve performance in the aircraft. This relationship between training in the simulator and performance in the aircraft is known as transfer of training.

Transfer of training can be thought of as the influence that experience or performance on one task has on some subsequent task.

There are three types of influences possible: (1) positive transfer, (2) negative transfer, and (3) zero transfer. Positive transfer occurs when prior task performance aids in performance on the subsequent task. Negative transfer exists when the prior task experience inhibits or disrupts the task that follows. Zero transfer is present if the prior task has no influence on the subsequent task (Ellis, 1965).

As applied to flight simulators, transfer of training is a measurement of how training in the simulator influences performance in the aircraft. Attempts are frequently made to measure transfer of training in terms of flight simulator hours required to replace aircraft hours. This relationship is usually expressed as a percentage of aircraft hours replaced by one simulator hour (LaRochelle, 1973). For example a 25 percent transfer of training would indicate that one hour of simulator time would equate to 15 minutes in the actual aircraft (Flexman, et al., 1954). There are numerous methods for computing this percentage. Each method produces a different percentage for the same simulator. This is illustrated in Table I. Five different expressions were used to compute the transfer of training percentage from the data collected during an actual evaluation of a training device. Even though the same data were used for each computation and each expression has been used in at least one official effectiveness report, a tremendous range of

Table I
Differences in Transfer of Training Expressions

Expression:	1	2	3	4	5
Transfer:	92%	71%	60%	12%	-54%

(From "Measures for the Efficiency of Simulators as Training Devices," 1967) transfer percentages resulted ("Measures for the Efficiency of Simulators as Training Devices," 1967). An invalid transfer of training percentage can be computed to fit the desired results by carefully choosing the computational method used (LaRochelle, 1973).

The technique most frequently used to measure transfer of training is the performance rating approach discussed in the Training Efficiency section of this chapter. The only difference is that performance is measured in the aircraft instead of in the simulator. This measurement approach does not separate the efficiency of original learning from the transfer of training. The validity of the approach is limited by the problems associated with the performance measurement approach. The limitations of the environment in an initial flight simulator evaluation again prohibit the use of this measurement technique.

This writer is of the opinion that a carefully constructed subjective evaluation of the simulator by instructor pilots can also provide a valid estimate of the transfer of training from the simulator to the aircraft. An experienced instructor pilot has observed the influence of many types of ground training on performance in the aircraft. This experience qualifies the instructor to make judgements about the influence of flight simulator characteristics on performance in the aircraft. Chapter IV discusses this subjective evaluation technique.

Fidelity of Simulation

A hypothesis common to the flight simulator industry is that the greater the degree of physical similarity between stimuli and responses in the simulator and those in the aircraft, the greater the amount of positive transfer of training that will take place (Ellis, 1965). This concept is based on theories such as the Osgood's Transfer Surface illustrated in Figure 3-2. The amount of transfer is indicated on the vertical axis, the degree of similarity of responses on the x axis, and the degree of similarity of stimuli on the y axis. The surface implies that the higher the degree of physical similarlity in both stimuli and responses, the higher positive transfer of training experienced. As either stimuli or responses decrease in physical similarity, the transfer of training decreases (Travers, 1963).

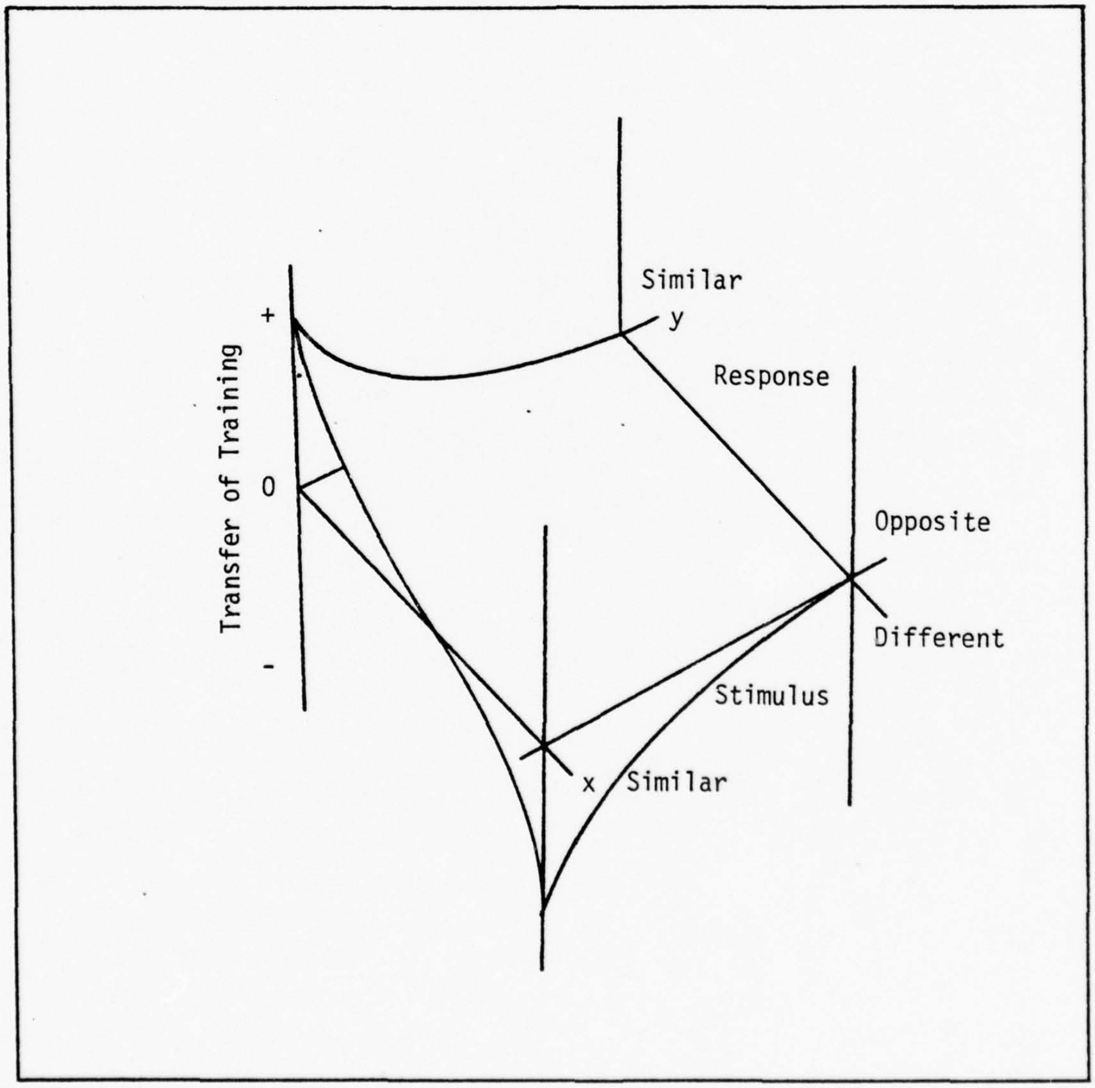


Figure 3-2. Osgood's Transfer Surface (From Ellis, 1965)

Even though this theory has been challenged in recent years by many researchers in the flight simulator and transfer of training areas, almost every current flight simulator evaluation technique reviewed in Chapter II concentrated on the fidelity of engineering simulation (Smode, Hall, & Meyer, 1966; Williges, et al., 1972; Harris, 1975). Using fidelity of engineering simulation as a criterion variable assumes that the highest quality of training and largest reduction in aircraft flying time are realized when flight in the simulator is not perceptably different from flight in the aircraft. Some of the reasons for this assumption are indicated by the following quote:

> The ultimate objective in developing a trainer is to build a device capable of training a person to a high level of proficiency and having a positive transfer of training from the simulator to the operational equipment it represents. The terms "high level of proficiency" and "positive transfer of training" are difficult to quantify and to directly relate to the performance characteristics necessary for inclusion into the simulator. Since these terms are difficult to quantify, the tendency is to include performance characteristics in the simulator if their necessity is doubtful resulting in the trainer having a higher cost than may be necessary. (Hood, 1975, p 369)

This is an excellent summary of the current attitude toward flight simulator development and evaluation. Training efficiency and transfer of training are extremely difficult to measure and to relate to flight simulator characteristics. This results in the emphasis being placed on fidelity of engineering simulation, which can be quantitatively measured and directly related to flight simulator characteristics.

It was stated earlier that simulator effectiveness is a function of the fidelity of engineering simulation. Many recent research efforts have been directed toward defining this relationship. Some examples of the findings of these research efforts are:

> The possibility exists that higher training value might be realized from trainers that do not "fly like" the aircraft in certain respects than from those that do. (Demaree, et al., 1964, p 2)

> There is considerable evidence, however, that deliberate deviations from fidelity of (engineering) simulation may lead to higher levels of transfer than does exact simulation. (Smode, Hall, & Meyer, 1966, p 167)

> Before any definite conclusions can be drawn about fidelity of (engineering) simulation, more detailed information is needed to determine how such variables as instructor ability, variations in the difficulty of the training task, and pilot experience level affect transfer performance. (Williges, et al., 1972, p 9)

> The question of what level of fidelity of flying qualities and performance is necessary for "effective" transfer of training is still an open issue.... (Harris, 1975, p 17)

These reports use the terms: training value, transfer, transfer performance, and effective transfer of training to refer to the effectiveness of the device. The general conclusion of each report is that a possibility exists that better simulator effectiveness might be realized with less than perfect fidelity of engineering simulation.

Fidelity of engineering simulation is the sole basis for the current Navy, Army, and Air Force evaluation techniques. However, recent research efforts indicate that simulator effectiveness is actually a nonlinear function of the fidelity of engineering simulation.

If we accept the hypothesis that simulator effectiveness can be maximized with less than perfect fidelity of engineering simulation, then simple cost considerations dictate that the fidelity of engineering simulation be less than perfect. A general relationship between cost, simulator effectiveness, and fidelity of engineering simulation is conceptualized in Figure 3-3. At 100% fidelity of engineering simulation, the flight simulator would have exactly the same physical characteristics as flight in actual aircraft represented. Therefore, the simulator effectiveness at 100% fidelity of engineering must be the same as the effectiveness of training in the actual aircraft. If the hypothesis is true that simulator effectiveness can be maximized at less than 100% fidelity of engineering simulation, then it must be possible to provide more effective training in the flight simulator than in the actual aircraft. Hence, we see the resultant simulator effectiveness curve rising above the actual aircraft effectiveness level in Figure 3-3.

Perhaps this concept can be clarified by an example. One of the most important pilot skills used in instrument flight is the ability to cross check the flight instruments well enough to interpret what the aircraft is doing and decide what control corrections need to be made. Generally, as the frequency of required control inputs increase, the required level of skill in the instrument cross check also increases. If the flight simulator requires the same stimulus-response relationship for instrument cross check as the aircraft, but more frequent control

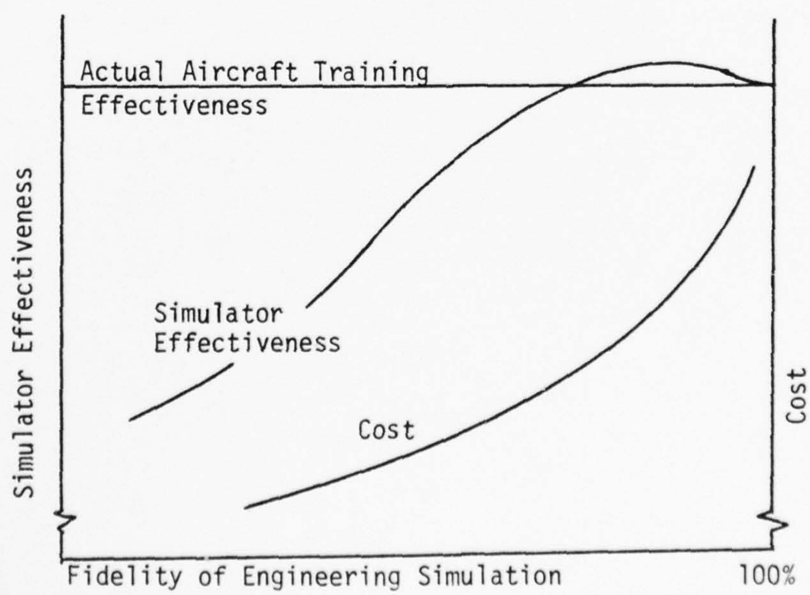


Figure 3-3. General Relationship of Simulator Effectiveness Cost, and Engineering Simulation (Adapted from Miller, 1954)

inputs than the aircraft, then the skill level necessary for instrument flight in the flight simulator will be higher than that required in the aircraft. As long as the stimulus-response relationship is the same in both situations, the flight simulator will provide more effective training than the aircraft.

Perfect, or 100%, fidelity of engineering simulation is not possible for all pilot tasks, though it is possible for simple tasks. The physical characteristics of the actual aircraft can be exactly duplicated in the flight simulator for simple tasks. The relationships for a simple task are illustrated in Figure 3-4. In this case, 100% fidelity of engineering simulation is possible at a cost c_1. However, maximum simulator effectiveness can be obtained at a cost of only c_2.

As the pilot tasks become more complex, exact or 100% fidelity of engineering simulation is not possible because of limitations in flight simulator technology. The relationships for a complex task are shown in Figure 3-5. Increasing complexity of the pilot task tends to shift the cost curve to the left. In this case, the most cost effective level of fidelity of engineering simulation is less than the maximum level possible within the technological limitation. It is somewhere in the range of diminishing returns shown on the illustration. The relationships for tasks of varying complexity would be somewhere in between the two cases used for illustration.

Unfortunately, the parameters of the curves used to illustrate this conceptualization of fidelity of engineering simulation and simulator effectiveness are not known. It is not possible to establish a level of fidelity of engineering simulation that would be the most cost effective for any given task. However, this conceptualization implies

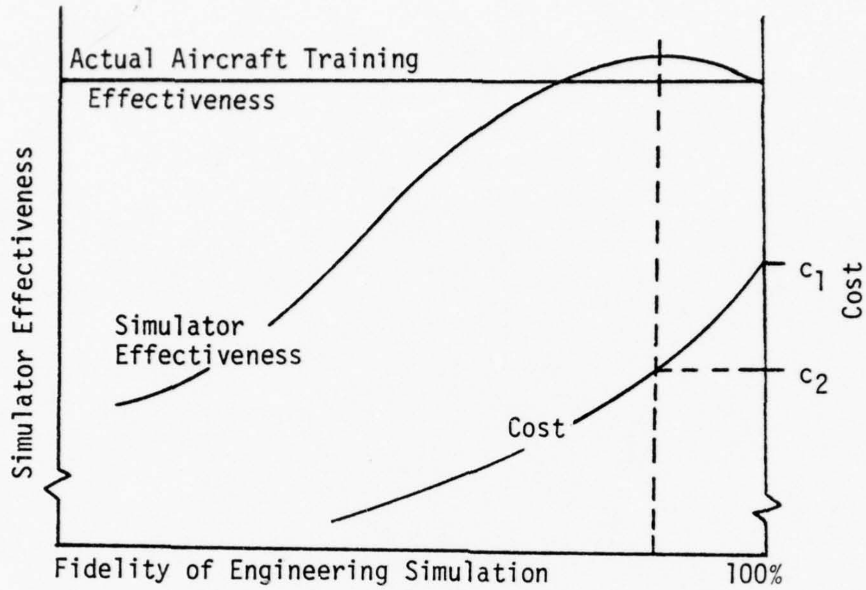


Figure 3-4. Simple Task Relationships

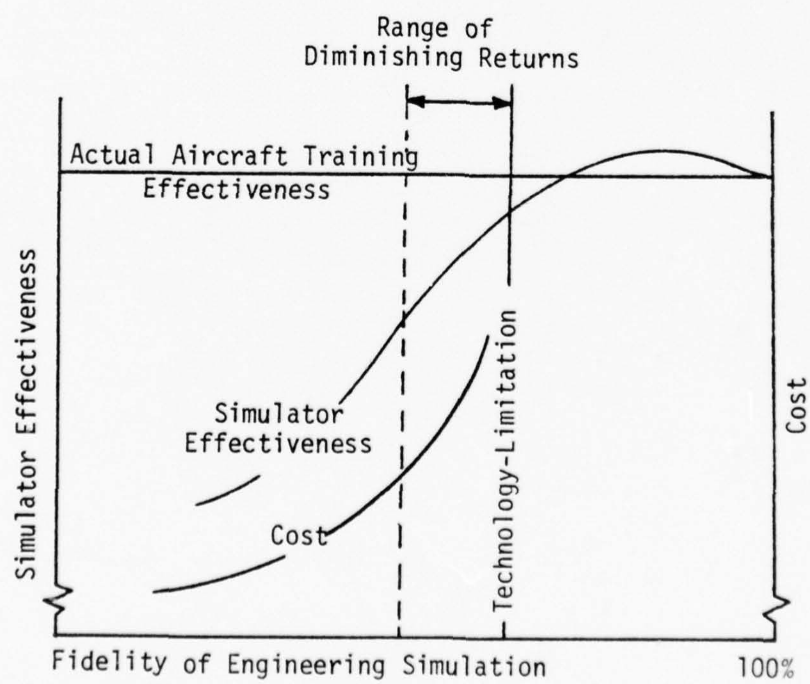


Figure 3-5. Complex Task Relationships

that a flight simulator could first be developed in the traditional quantitative manner to a lower level of fidelity of engineering simulation. This development could then be followed by an estimate of the simulator effectiveness for each task to be performed. Then, based on the estimates developed, the fidelity of engineering simulation could be increased for tasks which had not reached the point of diminishing returns in simulator effectiveness. The point of diminishing returns would have to be determined by a judgement based on the cost of increasing the fidelity of engineering simulation and the perceived increase in simulator effectiveness that would result.

Summary

The criterion variables measured during an initial flight simulator evaluation must provide adequate information on which to base a valid estimate of how well instrument and VFR flying skills can be taught and maintained in the flight simulator. Training efficiency, transfer of training, and fidelity of engineering simulation are the variables which appear to have the largest impact on the potential effectiveness of the simulator.

Training efficiency and transfer of training can be thought of as the key elements of fidelity of psychological simulation. Fidelity of pscyhological simulation is an expression of simulator effectiveness but cannot be measured with traditional performance rating evaluations due to the limitations on the initial flight simulator evaluation environment. In the opinion of the writer, a carefully constructed subjective evaluation of the flight simulator by instructor pilots, qualified in the aircraft being simulated, can produce a valid estimate of simulator effectiveness. This opinion is based on the assumption that experience with the teaching of flying skills qualifies instructor pilots to judge the impact of flight simulator characteristics on performance in the aircraft.

Fidelity of engineering simulation describes how well the physical qualities of flight in the real world aircraft have been copied. The current flight simulator development and evaluation techniques emphasize this variable. It can be conceptualized that the relationship between cost, simulator effectiveness, and fidelity of engineering simulation is dependent on the complexity of the task being simulated. This conceptualization implies that the most cost effective simulator design will have less than perfect fidelity of engineering simulation.

The criterion variable recommended for an initial flight simulator evaluation is simulator effectiveness, which is comprised of training efficiency and transfer of training. The initial flight simulator development should be conducted with the traditional qualitative approach to fidelity of engineering simulation. The resulting flight simulator characteristics should be refined by making the most cost effective improvements in fidelity of engineering simulation based on a subjective evaluation of simulator effectiveness.

IV. Flight Simulator Development and Evaluation Techniques

Introduction

The current flight simulator development and evaluation techniques discussed in Chapter II are based on fidelity of engineering simulation. The goal of these techniques is to produce a flight simulator which copies the physical characteristics of flight in the actual aircraft as accurately as possible within the budgetary and technological limitations of the program.

These techniques assume that simulator effectiveness is equivalent to the degree of fidelity of engineering simulation. However, the examination of criterion variables in Chapter III concluded that simulator effectiveness is not equal to, but merely a function of, fidelity of engineering simulation. Although the parameters of the function are not known, it has been conceptualized that as fidelity of engineering simulation is increased, a point is reached where diminishing returns in simulator effectiveness are experienced. Increasing fidelity of engineering simulation beyond this point, without prior knowledge of the estimated changes in simulator effectiveness, can commit funds to one area of fidelity of engineering simulation which might have been more effectively applied to another area.

The objective of the development and evaluation techniques recommended in this chapter is to develop an effective simulator through selective improvements in fidelity of engineering simulation. The recommended approach is a combination of quantitative and subjective techniques.

In this writer's opinion, traditional quantitative techniques should be used to design and develop the flight simulator to an initial minimum acceptable level of fidelity of engineering simulation. After this initial level is reached, subjective techniques should be used to select the areas of fidelity of engineering simulation in which additional improvements will yield the largest increase in simulator effectiveness.

Chapter Organization

The discussion of the development and evaluation techniques recommended by this effort will begin with the collection of data for construction of a mathematical model of the aircraft being simulated. Recommendations will then be made for the use of this model in a quantitative approach for establishing an initial level of fidelity of engineering simulation.

A subjective evaluation technique will then be introduced for the estimation of simulator effectiveness. The results of this subjective evaluation will provide the information necessary to select specific areas of fidelity of engineering simulation for additional improvements. The areas selected for additional improvements should be the ones which are estimated to yield the highest increase in simulator effectiveness when fidelity of engineering simulation is improved.

Quantitative Development to an Initial Level of Fidelity of Engineering Simulation

Design and initial development of a flight simulator must be based on a mathematical model which describes the characteristics of the aircraft that will be simulated. The current evaluation techniques reviewed in Chapter II indicate that there are normally two sources of data available for this model: (1) coefficients established during the development of the actual aircraft, and (2) data collected during flight tests conducted in the actual aircraft.

The coefficients established during the aircraft development are usually too inaccurate and incomplete for construction of a satisfactory model of the aircraft characteristics. These coefficients consist of estimates based on wind tunnel testing of the development aircraft components. Since the aircraft manufacturer is primarily interested in meeting performance specifications, these data are concentrated on the limits of the aircraft performance. Very little data are available from this source for modeling aircraft characteristics during normal operations. Therefore, flight tests must be conducted in the actual aircraft in order to complete the mathematical model.

The cost of the flight test program is directly related to the quantity and accuracy of the data collected during flight testing. The current development techniques discussed in Chapter II require the mathematical model to be very complete and accurate in order to develop a device with a maximum level of fidelity of engineering simulation. Therefore, the flight testing programs used during these development techniques were very extensive and costly.

The development and evaluation techniques recommended here do not require an extensive flight testing program. The mathematical model used for design and development need only be complete and accurate enough to establish an initial level of fidelity of engineering simulation. As a minimum, this initial level must be good enough to make the flight simulator controllable for each of the tasks that will be performed during the subjective evaluation.

In the opinion of this writer, an approach similar to the one used by the Naval Air Test Center for improvement of the mathematical model for Device 2F90 should be adequate for this purpose. The instrumentation need only consist of standard uncalibrated cockpit instruments and simple devices for the measurement of forces, distances, and time. The data can be collected in one aircraft or the measurements in several aircraft can be averaged.

The initial development coefficients and the flight test data will be combined for construction of an off-line model of the aircraft characteristics. This model must then be converted into a real-time mathematical model that can be programmed into the flight simulator computers. This process is illustrated in Figure 4-1.

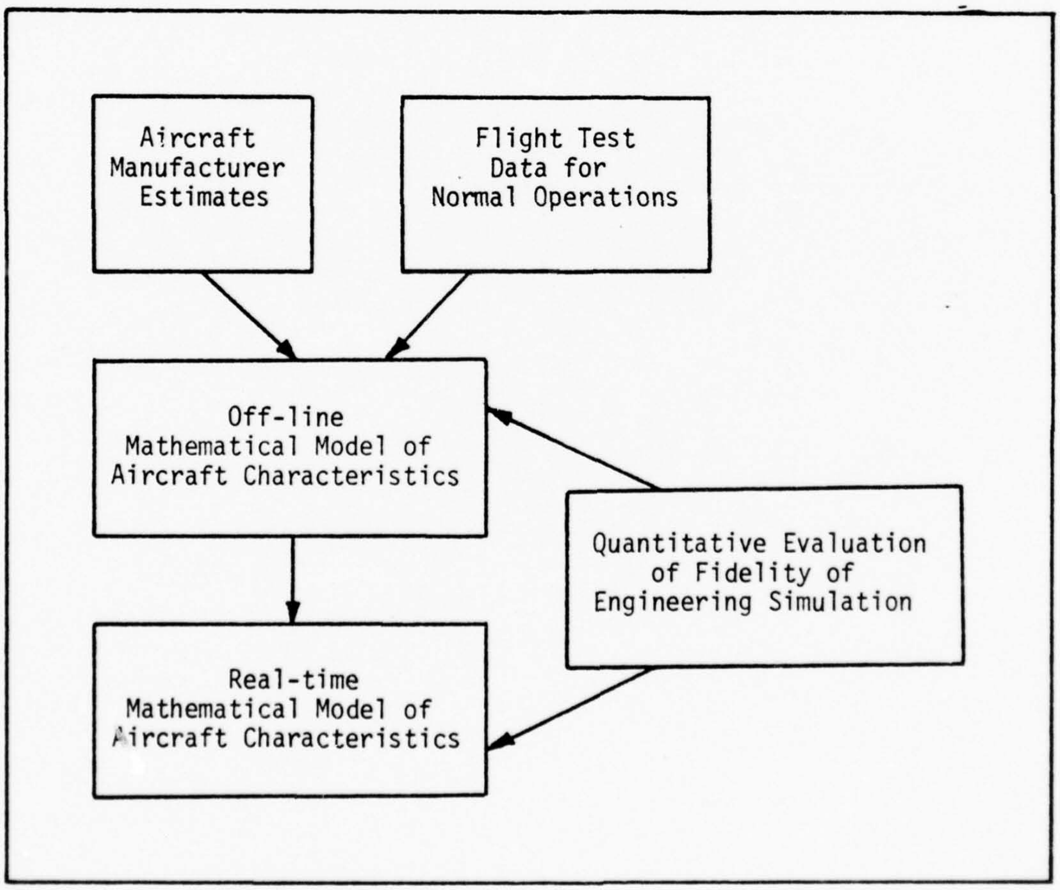


Figure 4-1. Construction of the Mathematical Model

The discussion of the UH-1H helicopter simulator, in Chapter II, indicated that some model accuracy is lost during the conversion from the off-line model to a real-time model. This loss of accuracy should be controlled through use of the traditional quantitative evaluation as illustrated in Figure 4-1. Tolerance limits should be established around the off-line model. The flight simulator characteristics should be measured and compared with the off-line model. Tradeoffs between the real-time model variables should then be made until the flight simulator characteristics are within the tolerances established around the off-line model.

In the writer's opinion, the application of these quantitative techniques will result in an adequate initial level of fidelity of engineering simulation. Additional improvements to the mathematical model associated with this initial level of fidelity of engineering simulation should not be made until the simulator effectiveness has been estimated for each pilot task to be performed in the flight simulator. These simulator effectiveness estimates will provide the necessary information to selectively improve the fidelity of engineering simulation and the mathematical model.

Purpose of the Subjective Evaluation

Subjective evaluation of flight simulators is not an original idea generated by this study. Qualified pilots were used for subjective evaluations in each of the Army, Navy, and Air Force techniques reviewed in Chapter II. However, this research effort does recommend a new purpose for the subjective evaluation techniques.

Subjective evaluations were used for the evaluation of Navy Device 2F90, Army Device 2B24, and the Air Force "tweaking" techniques. In each of these evaluations, the main purpose of the subjective evaluation was to identify deficiencies in fidelity of engineering simulation. The results of these evaluations were lists of flight simulator characteristics which the pilots perceived to be different than the actual aircraft characteristics.

The subjective evaluation recommended by this effort has two purposes: (1) identification of flight simulator characteristics that are perceived to be different than the actual aircraft, and (2) assessment of the impact that these differences will have on simulator effectiveness. The second purpose is the most important. The assessment of simulator effectiveness can provide the information required for selective improvements in fidelity of engineering simulation that will result in maximum simulator effectiveness.

The remaining sections of this chapter are devoted to the development of a subjective evaluation of flight simulator effectiveness. The aspects of a subjective evaluation discussed are: selection of participants, potential problem areas, an evaluation plan, and data collection methods.

Selection of Participants for the Subjective Evaluation

The participants selected for the subjective evaluation must be capable of identifying differences between the characteristics of the flight simulator and the real world aircraft. In addition, they must be able to assess the impact of these differences on flight simulator effectiveness.

Identification of differences requires extensive knowledge of the characteristics of each maneuver that will be taught in the simulator. Assessment of the impact of these differences on simulator effectiveness requires a detailed knowledge and understanding of the instructional techniques available for each maneuver, the average learning ability of students, and the most frequently encountered barriers to learning during each of the maneuvers.

The only personnel in the Air Force who meet these requirements are pilots who are currently qualified as instructors in the aircraft being simulated. The knowledge and understanding required of the subjective evaluation participants can only be obtained from the experience that comes with the teaching of flying skills. For this reason, it is recommended that all subjective evaluation participants be currently qualified as instructor pilots in the aircraft being simulated.

Some other factors which may adversely affect the accuracy of instructor pilot ratings should also be considered during the selection of participants. The factors discussed here are experience level and attitude toward the use of flight simulators to teach flying skills.

The experience level of an instructor pilot can be measured by both his total flying time and his flying time as an instructor pilot in the aircraft being simulated. The knowledge required for the evaluation of simulator effectiveness is primarily dependent on the pilot's total experience as an instructor pilot in the aircraft being simulated.

The impact that experience level has on the results of a subjective evaluation has not been defined as of this writing. For this reason, it can be argued that a broad range of experience levels should be included until an analysis of the results of a completed evaluation can determine the impact of experience levels on subjective evaluation ratings.

Very experienced instructor pilots have more knowledge on which to base their ratings. However, less experienced instructor pilots are usually more aware of the problems that they personally experienced while learning new flying skills. This research effort recommends that instructor pilots with varying degrees of experience be selected for the subjective evaluation. It also recommends extensive data analysis of a completed subjective evaluation to determine the impact of experience levels on the subjective evaluation ratings.

The instructor pilot's attitude toward the use of flight simulators could also affect the accuracy of his subjective ratings. Many Air Force pilots have a negative attitude toward flight simulation. This attitude is the result of bad experiences with current procedural trainers and the concern that flying time in the aircraft will be replaced by flight simulator time.

The relationship between an instructor pilot's attitude toward flight simulation and the accuracy of his ratings also has not been defined. Therefore, this research effort recommends that an analysis of the subjective ratings of instructor pilots with a positive, neutral, and negative attitude toward the use of flight simulation be conducted to determine the impact of pilot attitude on the accuracy of subjective ratings.

The recommended analysis of experience level and attitude toward flight simulation should provide the information required for the selection of participants in future subjective evaluations.

Potential Problem Areas in a Subjective Evaluation

There are several other factors within the evaluation environment which could also influence the accuracy of the simulator effectiveness estimates. The environmental factors which this researcher considers significant are: (1) rumors and the test atmosphere, (2) the use of partially operational equipment, and (3) changes to the real-time mathematical model. Each of these factors will be discussed in this section so that the subjective evaluation plan can be designed to minimize their impact on the accuracy of the evaluation results.

Rumors and the Evaluation Atmosphere. The evaluation atmosphere can be influenced by the attitudes of the subjective evaluation supervisors. Even if evaluation supervisors guard against making any direct statements about their perceptions of the simulator's capability, the overall attitude of the evaluation supervisors will be picked up by the participants performing the evaluation. For example, if the evaluation supervisors are generally disgusted with the simulator or the contractor and do not suppress this attitude during the evaluation, the participants could make overly critical evaluations. The opposite effect would occur if the evaluation supervisors were overly positive and did not suppress this attitude.

Rumors about specific deficiencies identified prior to the evaluation can also contaminate the data. If an instructor pilot has heard that a particular area of the flight simulator is weak, he will probably be overly critical of this area. Even more important, his concentration on one particular cue during a maneuver could prevent complete evaluation of other cues observed during the maneuver.

These types of problems are probably the hardest to avoid during a subjective evaluation. The evaluation supervisors must realize how easily they can influence the evaluation results. This influence can be partially controlled by minimizing the interaction between test supervisors and participants during the evaluation. For this reason, it is recommended that the participants fly the simulator solo during the evaluation and make their own evaluation with minimum assistance from the evaluation supervisors.

The Use of Partially Operational Equipment. Simulator effectiveness for a given maneuver will be dependent on each of the major cues presented in the flight simulator. Major cues are considered to be: visual, motion, aural, control forces, and instrument displays. If one or more of the major cues are not available due to inoperable equipment, the evaluation of simulator effectiveness could be contaminated by negative spillover effects into areas that are actually quite effective.

The cockpits to be used for evaluation of the T-37 Undergraduate Pilot Training Instrument Flight Simulator (UPT-IFS) are a good example of partially operational equipment. Two of the four cockpits to be used are not equippped with visual display systems. In the opinion of this writer, these two incomplete cockpits should not be used for the evaluation of simulator effectiveness.

Simulators can also become partially inoperable during the evaluation because of maintenance failures. In general, any failures which eliminate one or more of the major cues will prevent the collection of meaningful data because of possible spillover effects on the other major cues. It is recommended that in the event of such a failure, the evaluation be stopped until corrections can be made.

Changes to the Real-time Mathematical Model. It is very important that the parameters of the mathematical model remain constant throughout the evaluation. In order to arrive at an estimate of simulator effectiveness for a particular maneuver, all of the ratings assigned for that maneuver must be examined collectively. A change in the parameters of the mathematical model during the evaluation would make data collected prior to the change incompatible with data collected after the change.

The temptation to change a parameter during the evaluation when, for example, it appears that all evaluation participants will rate the cues related to that parameter low, must be resisted. If changes are considered necessary, they should be made early enough in the evaluation to allow adequate data for the estimation of simulator effectiveness to be collected after the change. If such a change occurs, the data collected prior to the change should be discarded or treated separately.

A Subjective Evaluation Plan

The evaluation plan presented in this section has been designed to produce an accurate estimate of simulator effectiveness. The design also attempts to minimize the impact of the potential problem areas discussed earlier. The recommended subjective evaluation plan consists of: (1) an initial briefing for the participants, (2) a simulator orientation flight for the participants, and (3) data collection during the evaluation. Recommend procedures for each of these areas are developed in the following sections.

Initial Briefing for Participants. In order to obtain the most useful data from the subjective evaluation, each instructor pilot should receive a thorough briefing prior to beginning the evaluation. This briefing should inform the participants of the importance of each of their inputs and the value of having ratings based on each individual's personal knowledge and experience. The procedures to be used for the evaluation should be described in detail to include a discussion of how the data will be used (Cooper & Harper, 1969).

The requirement for written comments to explain the ratings made should also be discussed. Emphasis should be placed on when comments are required and what information they should contain.

The evaluation participants should also be made aware of how perceived attitudes toward the simulator and preconceived evaluations based on rumors can adversely affect the evaluation results. This portion of the briefing should explain how concentration on one or more of the major cues can result in an inaccurate rating for other major cues.

The briefing must emphasize the fact that an estimate of simulator effectiveness is the desired result of the evaluation. Observations made during an experimental subjective evaluation of the T-37 UPT-IFS indicated that a subjective evaluation participant tends to concentrate on deficiencies in fidelity of engineering simulation rather than on simulator effectiveness. This point must be emphasized during the briefing.

Simulator Orientation Flight. One of the recommendations made earlier was to minimize the interactions between evaluation participants and supervisors during the rating process. In the writer's opinion, the best way to minimize interactions is to have each evaluation participant fly and rate the simulator by himself without the assistance of evaluation supervisors. In order for this to be done, the evaluation participants must learn the operation of controls which are unique to the flight simulator. A short orientation flight is recommended for this familiarization with unique control features.

It is imperative that the evaluation supervisor and participant do not discuss the characteristics of the flight simulator during this orientation flight. The evaluation supervisor must not make any statements which might influence the participant's ratings during the evaluation which will follow.

Data Collection during the Evaluation. The initial briefing and orientation flight should adequately prepare the participants for the subjective evaluation. During the evaluation, each participant should rate the simulator effectiveness for each maneuver that will be taught during operational use of the flight simulator. These maneuvers should be flown as many times as necessary for an accurate rating of each of the cues presented during the simulation.

The problem of a pilot quickly adapting to the simulator characteristics, which was identified by the Army and Navy evaluations discussed in Chapter II, should not be a significant factor during this evaluation. It is intended that instructor pilots are temporarily brought in from the field to participate in the evaluation. These participants should have a flight in the real world aircraft just before the evaluation and should not be expected to participate more than a few days in the simulator evaluation.

The ratings awarded during the evaluation should be recorded on a questionnaire. Part of the questionnaire developed for the subjective evaluation of the T-37 UPT-IFS is included in Appendix A of this report as an example of the recommended subjective evaluation questionnaire. The simulator effectiveness is rated separately for each maneuver performed. In addition, the major cues which are perceived to be different from the actual aircraft cues are identified. If the simulator is given a low effectiveness rating, additional comments must be made to identify the major cue or cues which caused the low rating and the exact nature of the deficiency in the major cue or cues involved.

The rating scale included in the T-37 UPT-IFS questionnaire is a modification of one developed by this effort. The purpose of this rating scale was to improve the consistency of the results and to insure that evaluation participants evaluated simulator effectiveness instead of simply fidelity of engineering simulation.

The rating scale originally used by Air Training Command for this questionnaire was:

5. Simulator performance exactly duplicates the aircraft.

4. Minor differences between simulator and aircraft performance were noted, but would not detract from the training capabilities of the simulator.

3. Minor differences between simulator and aircraft performance were noted, which could have a minor impact on the training capabilities of the simulator.

2. Major differences between simulator and aircraft performance were noted, which could impact on the training capabilities of the simulator.

1. Major differences between simulator and aircraft performance were noted, which could have a major impact on the training capabilities of the simulator. (UPT-IFS Integrated Test, May 1976)

This type of rating scale presents two major problems. The evaluation participant would probably have a great deal of difficulty remembering the meaning of each rating during the evaluation. Even if the meanings were reviewed during each rating, the lack of concrete definitions for such terms as "minor differences" and "major differences" or "could impact" and "could have a major impact" would surely produce different ratings depending on how each individual participant perceived the terms.

It would also be extremely difficult to base fidelity of engineering simulation improvements on this scale. It is almost impossible to say which rating would require improvements regardless of the cost and time involved, which rating would require improvements only if the cost and time involved were considered to be reasonable, and which ratings do not indicate a need for improvements in fidelity of engineering simulation.

In order to simplify the evaluation process and save time during the evaluation, the original rating scale was changed to the following:

1. Simulator matches aircraft.
2. Simulator different than aircraft - No impact on training capabilities.
3. Simulator different than aircraft - Degrades training capabilities. (UPT-IFS Integrated Test, May 1976)

This scale eliminates the problems of complexity and undefined terms in the original scale. However, it would provide data that was even less useful for improving selection decisions.

The writer recommends a five-point rating scale with simple definitions. The recommended scale is:

5. Simulator exactly duplicates the aircraft.
4. Deviations exist, but will not affect the training capabilities.
3. Deviations exist, but will have an insignificant effect on the training capabilities.
2. Deviations exist and will significantly affect the training capabilities, but will not prevent using the simulator to teach this maneuver.
1. Deviations exist which prevent using the simulator to teach this maneuver.

This scale limits the definitional problem to the terms "significant" and "insignificant" effect. It also provides a good basis for selective improvement decisions. A rating of "1" requires fidelity of engineering simulation improvements at any cost if the maneuver is to be taught in the simulator. A rating of "2" requires improvements if the costs are considered to be reasonable. Ratings "3" through "5" do not require improvements in fidelity of engineering simulation but should provide valuable information for development of the simulator training program.

Regardless of the rating scale used, one serious problem will affect the resulting data. This problem is created by attaching a firm meaning to each rating and expecting the participant to remember the meaning of the rating throughout the evaluation. Also, this problem is further complicated by possible subjective variations in each participant's perception of these meanings. Observations during an experimental subjective evaluation of the T-37 UPT-IFS clearly showed that the participant reverted to a pure "good-fair-poor" subjective scale after a short time. This is, in the case of the three-point scale used, the rater viewed a "1" as the best performance possible, a "3" as some level of poor performance, and a "2" as average. It was also observed that the concept of assessing the impact on simulator effectiveness was quickly disregarded. The ratings given represented only the degree of realism of the simulator.

It is believed that this problem can be almost completely solved through the use of a sequence of dichotomous decisions similar to the one used by the Air Force Test and Evaluation Center (AFTEC) for pilot evaluation of aircraft handling characteristics. The rating technique used by AFTEC is commonly known as the Cooper-Harper Scale (Cooper & Harper, 1969). The modified version of this decision sequence recommended for use in a subjective evaluation of flight simulator effectiveness is shown in Figure 4-2. The participant uses this rating technique by answering each of the four questions yes or no for each area rated. A "no" answer results in an immediate rating assignment while a "yes" answer requires one or more of the following questions in the sequence to be answered. It is believed that this technique will produce consistent and useful results.

The main objection raised against this technique is that a great deal of time is required to follow the decision sequence for each rating. However, it is felt that the value of consistent and useful data far outweighs the additional time required. Furthermore, a few practice sessions with this decision sequence chart on a few maneuvers should quickly internalize the chart, so that the participant could arrive at a rating in a reasonable time. In the extreme case, it would still be

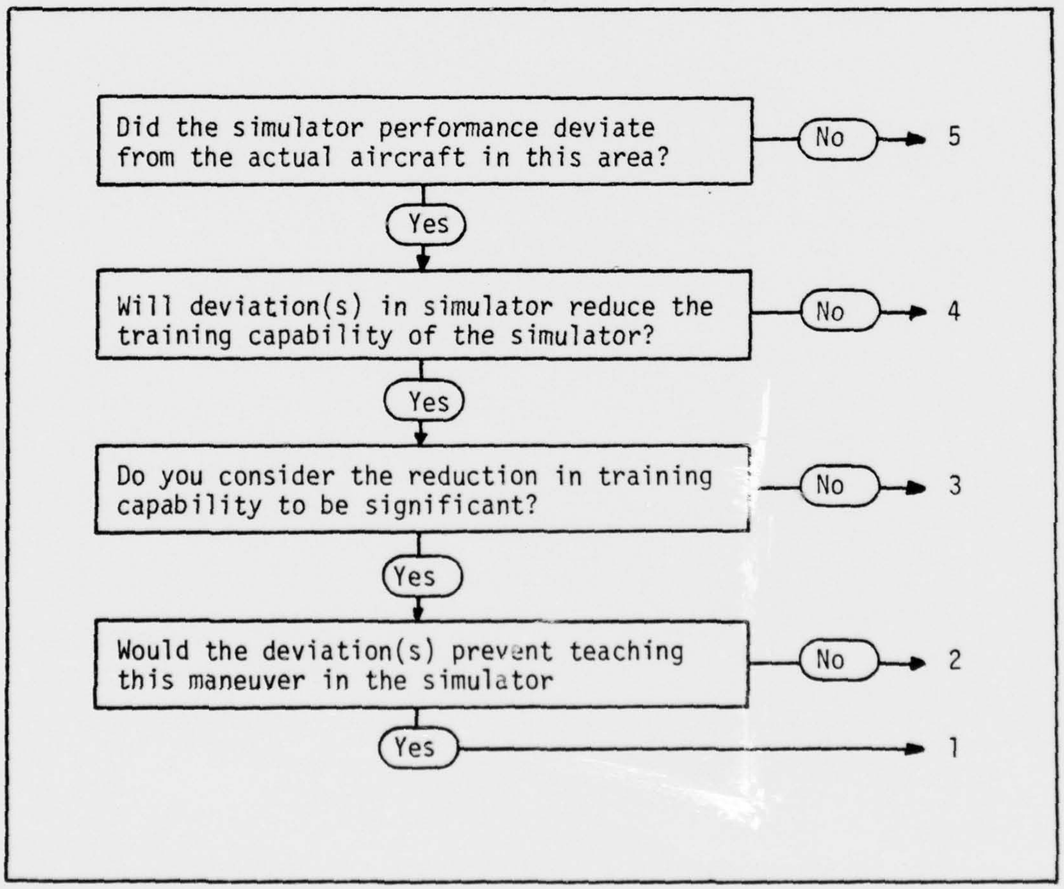


Figure 4-2. Recommended Dichotomous Decision Sequence

better to have a small amount of consistent and accurate data than a large amount of purely subjective data.

The rating scale developed above was adapted by Simulator System Program Office and Air Training Command personnel and included in the T-37 UPT-IFS questionnaire in the form shown in Appendix A.

Summary

The flight simulator development and evaluation techniques recommended by this research effort consist of quantitative and subjective evaluations. In the opinion of this writer, traditional quantitative techniques should be used to develop an initial, flyable level of fidelity of engineering simulation. Subjective evaluations of simulator effectiveness should then be used to select the areas of fidelity of engineering simulation in which improvements would yield the greatest increase in simulator effectiveness.

It is recommended that only currently qualified instructor pilots be used as participants in the subjective evaluation. The selected participants should represent a good cross section of experience and attitudes toward the use of flight simulators until further research can assess the impact of these factors on the results of the subjective evaluation.

The subjective evaluation plan recommended by this research effort uses an initial briefing and an orientation flight in the simulator to prepare each participant for the evaluation. This preparation allows the interactions between evaluation supervisors and participants to be minimized during the actual rating process.

A dichotomous decision process is recommended as an aid for assigning subjective simulator effectiveness ratings. The rating scale and questionnaire developed for the T-37 UPT-IFS subjective evaluation are included in Appendix A as an example of the subjective evaluation technique recommended by this effort.

V. Summary and Recommendations

The primary objectives of this research effort were to: (1) determine which criterion variables are most relevant to an initial evaluation of flight simulator effectiveness, and (2) develop a technique which could be used for the evaluation of these criterion variables. This chapter summarizes the examination of criterion variables and the recommended techniques for flight simulator development and evaluation. The final section recommends an area for further research.

Criterion Variables for Evaluation

The determination of applicable criterion variables was based on a review of current flight simulator evaluation techniques and related literature. The criterion variable stressed by most current flight simulator evaluation techniques is fidelity of engineering simulation, which is a measurement of how well the physical characteristics of flight in the real world aircraft have been copied by the flight simulator.

The examination of the Air Force flight simulator mission and desired quality for an Air Force flight simulator, in Chapter II of this report, identified simulator effectiveness as the most desirable criterion variable. Simulator effectiveness is how well the flight simulator will perform its intended mission of reducing aircraft flying time used for training by providing an effective environment in which to teach and maintain flying skills in instrument and visual flight rule conditions.

The conceptualized relationships between the criterion variables examined can be summarized by the following set of equations:

$$
\begin{aligned}
\text{Simulator Effectiveness} &\equiv \text{Aircraft Flight Time Saved} + \text{Quality of Training} \\
&\equiv \text{Training Efficiency} + \text{Transfer of Training} \\
&\equiv \text{Fidelity of Psychological Simulation} \\
&\equiv f\ (\text{Fidelity of Engineering Simulation})
\end{aligned}
$$

The major finding of the examination of criterion variables was a conceptualization of the relationships between fidelity of engineering simulation, simulator effectiveness, and the cost of the simulator. Although the parameters of this relationship are not known, it has been conceptualized that as fidelity of engineering simulation is increased, a point is reached where diminishing returns in simulator effectiveness are experienced. In the opinion of this writer, increasing fidelity of engineering simualtion beyond this point, without prior knowledge of the estimated changes in simulator effectiveness, can be an uneconomical approach to simulator development.

Recommended Flight Simulator Development and Initial Evaluation Techniques

The techniques recommended for development and initial evaluation of a flight simulator are based on the review of current techniques and related literature, observations during the evaluation of the T-37 Undergraduate Pilot Training Instrument Flight Simulator (UPT-IFS), numerous informal interviews with Simulator System Program Office and Air Training Command personnel, and four years of personal experience as an instructor pilot in Undergraduate Pilot Training.

The recommended techniques consist of a combination of the traditional qualitative techniques plus some subjective techniques. This writer recommends that a flight simulator first be designed and developed to an initial minimum acceptable level of fidelity of engineering simulation through the use of the traditional quantitative techniques. The design and initial development process will require construction of a mathematical model which represents the characteristics of flight in the aircraft being simulated. The sources of data available for construction of this model are the aircraft manufacturer's estimates and flight test data. This researcher recommends an inexpensive flight test program which uses standard cockpit instrumentation and simple devices for the measurement of forces, distances, and time. Quantitative evaluation techniques are recommended to assure that the simulator flight characteristics adequately represent the mathematical model of the aircraft flight characteristics.

Following development to an initial minimum acceptable level of fidelity of engineering simulation, a subjective evaluation of simulator effectiveness is recommended. The subjective evaluation used should provide the information necessary for the selection of areas of fidelity of engineering simulation in which improvements would yield the greatest increase in simulator effectiveness.

The participants selected for the subjective evaluation must be capable of identifying differences between the flight characteristics of the simulator and the real world aircraft and be able to assess the impact of these differences on flight simulator effectiveness. The writer is of the opinion that the only Air Force personnel who meet these requirements are currently qualified instructor pilots in the aircraft being simulated. For this reason, it is recommended that all subjective evaluation participants be currently qualified as instructor pilots in the aircraft being simulated.

The recommended subjective evaluation plan uses an initial briefing and an orientation flight to prepare each participant for the evaluation and to limit the impact of factors which may influence the accuracy of the results. This preparation allows the interactions between the evaluation supervisors and participants to be minimized during the actual rating process.

A dichotomous decision sequence has been developed and is recommended as an aid for assigning subjective evaluation ratings. The use of this decision sequence should improve the consistency and accuracy of the subjective evaluation results. The rating scale, decision sequence, and questionnaire developed for the T-37 UPT-IFS subjective evaluation are included in Appendix A of this report as an example of the subjective evaluation technique recommended by this research effort.

Recommendations for Further Research

Additional research is needed to identify and measure the factors which influence the accuracy of an instructor pilot's subjective evaluation of simulator effectiveness. It is recommended that the results of a completed subjective evaluation be analyzed to determine, at a minimum, how experience level and attitude toward flight simulation impact on the accuracy of the results. This research should provide the information required for more effective selection of participants in future subjective evaluations.

BIBLIOGRAPHY

1. AFR 80-14. Research and Development - Test and Evaluation. Washington: Department of the Air Force, February 1975.

2. Caro, Paul W. "Aircraft Simulators and Pilot Training." Human Factors, Vol. 15, No. 6:pages 502-509 (December 1973).

3. Caro, Paul W., Robert N. Isley, and Oran B. Jolly. Mission Suitability Testing of an Aircraft Simulator. Technical Report 75-12. Alexandria, Virginia: Human Resources Research Organization, June 1975.

4. Catron, Robert L. A New Approach for Establishing Aerodynamic Performance of Flight Trainers. 8th NTEC/Industry Conference Proceedings. Orlando, Florida: Naval Training Equipment Center, November 1975.

5. Cooper, George E. and Robert P. Harper, Jr. The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities. NASA Technical Note D-5153. Washington: National Aeronautics and Space Administration, April 1969.

6. Demaree, Robert G., Don A. Norman, and William G. Matheny. An Experimental Program for Relating Transfer of Training to Pilot Performance and Degree of Simulation. Technical Report. Fort Worth, Texas: Life Sciences, Inc., July 1974.

7. Dunlap, Duane S., et al. Air Force Master Plan - Simulators for Aircrew Training. Final Report. Wright-Patterson Air Force Base, Ohio: Aeronautical Systems Division, December 1975.

8. Ellis, H. C. The Transfer of Learning. New York: The MacMillan Company, 1965.

9. Flexman, Ralph E., John C. Townsend, and George N. Ornstein. Evaluation of a Contact Flight Simulation When Used in an Air Force Primary Pilot Training Program: Part I. Overall Effectiveness. Technical Report. Goodfellow Air Force Base, Texas: Air Force Personnel and Training Research Center, September 1954.

10. Fogarity, Laurence E. Survey of Flight Simulation Computation Methods. SAE conference proceedings. New York: Society of Automotive Engineers, Inc., April 1967.

11. Galloway, R. T., Jr. and M. D. Hewett. Flight Fidelity--Improvement Program for the TA-4J Operational Flight Trainer (Device 2F90). Final Report Patuxent River, Maryland: Naval Air Test Center, March 1975.

12. Harris, William T. Device 2F90 Flying Qualities and Performance Evaluation and Discrepancy Correction. Technical Report. Orlando, Florida: Naval Training Equipment Center, September 1975.

13. Hewett, M. D. and R. T. Galloway. Flight Fidelity Validation for the TA-4J Operational Flight Trainer (Device 2F90). Report of Test Results. Patuxent River, Maryland: Naval Air Test Center, September 1975.

14. Hewett, M. D., R. T. Galloway, and J. C. Murray. On Improving the Flight Fidelity of Operational Flight/Weapons System Trainers. 7th NTEC/Industrial Conference Proceedings. Orlando, Florida: Naval Training Equipment Center, November 1974.

15. Heinle, Leon C. "Surrogate Flying." Air Force Magazine, Vol. 56, No. 4: pages 59-61 (April 1973).

16. Hood, James D. Simulation Cost Versus Fidelity. 8th NTEC/Industrial Conference Proceedings. Orlando, Florida: Naval Training Equipment Center, November 1975.

17. LaRochelle, Donald Z. A Cost Effective Analysis of an Air Combat Simulator. Unpublished thesis. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, June 1973.

18. Lewis, Oliver W. "Simulation--The New Approach." Air University Review, Vol. 25, No. 3: pages 41-55 (March-April 1974).

19. "Measures for the Efficiency of Simulators as Training Devices." Ergonomics, Vol. 10, No. 1: pages 63-65 (1967).

20. Miller, Robert B. Psychological Considerations in the Design of Training Equipment. Technical Report. Wright-Patterson Air Force Base, Ohio: Wright Air Development Center, December 1954.

21. Rhodes, John. The Urgent Need for Flight Simulators for Present and Future Aircraft. SAE conference proceedings. New York: Society of Automotive Engineers, Inc., April 1967.

22. Richmond, James A. Verisimilitude Testing: A New Approach to the Flight Testing of Air Force Simulators. Conference paper. Wright-Patterson Air Force Base, Ohio: Aeronautical Systems Division, August 1976.

23. Rust, Steven K. Flight Simulator Fidelity Assurance. 8th NTEC/Industry Conference Proceedings. Orlando, Florida: Naval Training Equipment Center, November 1975.

24. Smode, Alfred F., Eugene R. Hall, and Donald E. Meyer. An Assessment of Research Relevant to Pilot Training. Technical Report. Wright-Patterson Air Force Base, Ohio: Aerospace Medical Research Laboratories, November 1966.

25. Snow, Jackie L. Simulation in Undergraduate Pilot Training: The Practical Road to Specialization. Professional study. Maxwell Air Force Base, Alabama: Air War College, April 1975.

26. *Training Devices and Simulators: Some Research Issues*. Technical Report AFPTRC 54-16, May 1954.

27. Travers, R. M. W. *Essentials of Learning*. New York: The Macmillan Company, 1963.

28. *UPT-IFS Integrated Test*. Test questionnaire. Randolph Air Force Base, Texas: Air Training Command, May 1976.

29. Walker, L. A. and R. T. Galloway, Jr. *Flight Fidelity Evaluation of the T-2C Operational Flight Trainer*, Device 2F101. Final Report. Orlando, Florida: Naval Air Test Center, January 1975.

30. *Webster's Seventh New Collegiate Dictionary*. Massachusetts: G&C Merriam Company, 1963.

31. Williges, Deverly H., *et al*. *Synthetic Flight Training Revisited*. Technical Report ARL-72-21/AFOSR-72-10. Savoy, Illinois: University of Illinois, August 1972.

RELATED REFERENCES

Caro, Paul W., Robert N. Isley, and Oran B. Jolly. The Captive Helicopter as a Training Device: Experimental Evaluation of a Concept. Technical Report. Fort Rucker, Alabama: Human Resources Research Office, June 1968.

Chalke, G. P. and R. T. Galloway. Flight Fidelity Evaluation of the A-7E Weapons Systems Trainer, Device 2F111. Second Interim Report. Patuxent River, Maryland: Naval Air Test Center, June 1975.

Chapman, G. Courtney. An Experimental Assessment of a Ground Pilot Training in General Aviation. Technical Report. Columbus, Ohio: Ohio State University, February 1966.

Crook, Warren G. Experimental Assessment of Ground Trainers in General Avaiation Pilot Training. Final Report. Washington: Federal Aviation Administration, April 1967.

Curtis, E. R., P. A. Day, and W. A. Allison. NADC Air-to-Air Combat Simulator Evaluation. Third Interim Report. Patuxent River, Maryland: Naval Air Test Center, August 1971.

Dougherty, Dora J., Robert C. Houston, and Douglass R. Nichlas. Transfer of Training in Flight Procedures from Selected Ground Training Devices to the Aircraft. Technical Report NAVTRADEVCEN 71-16-16. Urbana, Illinois: University of Illinois, September 1957.

Duncan, Carl P. and Benton J. Underwood. The Effect on Transfer of Varying Stimulation During Training. WADC Technical Report 56-279. Wright-Patterson Air Force Base, Ohio: Wright Air Development Center, December 1957.

Ewing, K. W., R. T. Galloway, and R. L. McKinney. Evaluation of the A-7E Night Carrier Landing Trainer, Device 2F103. Final Report. Patuxent River, Maryland: Naval Air Test Center, February 1975.

First Article Qualification Test Procedure - T-37 Flight Simulator. Acceptance Test Procedure. Binghamton, New York: Singer Simulation Products, May 1976.

Geiselhart, Richard, Paul T. Kemmerling, Jr., and Joe Y. Yasutake. Evaluation of the Terrain Avoidance Function of a Multimode Radar Display. Final Report. Wright-Patterson Air Force Base, Ohio: Aeronautical Systems Division, June 1974.

Hickey, Leo F., Wayne E. Springer, and Francis L. Cundari. A Development in Cockpit Geometry Evaluation. Research Report. Seattle, Washington: Boeing Company, November 1968.

Isley, Robert N., Paul W. Caro, Jr., and Oran B. Jolly. Evaluation of Synthetic Instrument Flight Training in the Officer/Warrent Officer Rotary Wing Aviation Course. Technical Report 68-14. Fort Rucker, Alabama: Human Resources Research Office, November 1968.

Jeantheau, Gabriel G. Handbook for Training Systems Evaluation. Technical Report. Orlando, Florida: Naval Training Device Center, January 1971.

Keene, G. F. and L. A. Lantcer. NADC Air Combat Simulator Evaluation. Second Interim Report. Patuxent River, Maryland: Naval Air Test Center, November 1968.

Koch, Carl and D. J. Dougherty. Contact Analog Simulator Evaluations - Unusual Attitude Maneuvers. Technical Report. Fort Worth, Texas: Bell Helicopter Company, December 1965.

Mahler, W. R. and G. K. Bennett. Special Devices in Primary Flight Training: Their Training and Selection Value. A Summary Report. Springfield, Virginia: National Technical Information Service, August 1949.

Mahler, W. R. and G. K. Bennett. Psychological Studies of Advanced Naval Air Training: Evaluation of Operational Flight Trainers. Technical Report. New York: The Psychological Corporation, September 1950.

Mendenhall, William. The Design and Analysis of Experiments. Belmont, California: Wadsworth Publishing Company, 1968.

Michaels, Jerome L. Study and Evaluation of V/STOL Ground-Based Simulation Techniques for the X-22A Aircraft. USAAVLABS Technical Report 69-40. Buffalo, New York: Bell Aerosystems Company, June 1969.

O'Gorman, K. P., et al. An Evaluation of the Operational Capabilities and Fuel Conservation Implications of Using Remotely Piloted Vehicles, Advanced Aircraft Flight Simulators, and Lighter-Than-Air Aircraft. Unpublished thesis. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, August 1975.

Phillips, C. R., Jr. An Experimental Assessment of a Ground Pilot Trainer in General Aviation. Technical Report. Miami, Florida: Miami-Dade Junior College, January 1966.

Ripley, Roy L. Undergraduate Pilot Training: Decisions for the Future. Professional Study. Maxwell Air Force Base, Alabama: Air War College, April 1975.

Spahr, B., D. Bischoff, and R. vonHusen. Flight Fidelity Evaluation of KC-130F Operational Flight Trainer, Device 2F107. First Interim Report. Patuxent River, Maryland: Naval Air Test Center, May 1975.

Test Plan--Undergraduate Pilot Training Instrument Flight Simulator (UPT-IFS) Operational Test and Evaluation. Test Plan. Randolph Air Force Base, Texas: Air Training Command, December 1975.

Townsend, John C. Evaluation of the Link, ME-1, Basic Instrument Flight Trainer. Development Report. Randolph Air Force Base, Texas: Air Force Personnel and Training Research Center, June 1956.

Westbrook, C. B. *Background of Piloted Simulator Development*. Technical Report. Wright-Patterson Air Force Base, Ohio: Air Force Flight Dynamics Laboratory, August 1964.

Willich, W. and R. E. Edwards. *Analysis and Flight Simulator Evaluation of an Advanced Fighter Cockpit Configuration*. Technical Report. Seattle, Washington: Research and Engineering Division, Boeing Aerospace Company, March 1975. AD A009900.

Zamarin, D. M. and D. I. Blom. *Simulator Evaluation of Five Counter-Pointer Altimeters*. Technical Report. Long Beach, California: Douglas Aircraft Company, October 1970.

APPENDIX A

T-37 UPT-IFS RATING SCALE AND QUESTIONNAIRE

C O P Y

UNDERGRADUATE
PILOT TRAINING
INSTRUMENT FLIGHT
SIMULATOR
(UPT-IFS)

PRELIMINARY FIRST ARTICLE
QUALIFICATION TESTING

T-37
INTEGRATED TEST - ANNEX D

BOOK 1D OF 3
MANEUVER EVALUATION GUIDES
OPERATIONS

MEG # 1

PREPARED BY
SIMULATOR MGT BRANCH
3305TH SCHOOL SQUADRON

PREPARED IAW ASD
EXHIBIT ENCT-73-1
SECTION 3

MAY 1976

COPY

IP Maneuver Evaluation Guide #1

INSTRUCTIONS: This maneuver evaluation guide is designed to evaluate simulator fidelity for each of the maneuvers listed in the guide. It is not representative of an actual training mission; therefore, it is not desired that you make radio calls or attempt to exercise the console operator in the role of an air traffic controller.

Some operator assistance is required. Prior to your entering the cockpit, you should inform the operator that you are flying maneuver evaluation guide #1 so that he can set in the appropriate initial conditions.

The initial conditions for this mission are: Runway 19 at Bristol Lake AFB, VOR/DME 109.8, channel 35. The local altimeter is 29.92. Ceiling and RVR are set at 600' and 5000' respectively. You will break out of the clouds at 8000' MSL during the climb.

After entering the cockpit and strapping in, insure that the cockpit configuration is correct for takeoff, i.e. engines running, flaps 50%, gear down. Then call the operator over UHF and request motion consent and initial condition insertion. Adjust aircraft sound volume to the appropriate level. Prior to starting thru this guide, you will have to unfreeze the simulator.

C O P Y

MANEUVER DESCRIPTION	RATING SCALE	Circle any of the areas below in which differences existed between the simulator and the aircraft.	COMMENTS
Takeoff (FREEZE after retracting gear and flaps)	RATING SCALE Use the following scale to compare the effectiveness of the simulator and the aircraft as training devices. The simulator would be a more effective training device than the aircraft. — YES → 5 NO ↓ The simulator would be just as effective as the aircraft. — YES → 4 NO ↓ The simulator would be less effective than the aircraft, but could be used for a major portion of the training. — YES → 3 NO ↓ The simulator would be less effective than the aircraft, but could be used for some portion of the training. — YES → 2 NO → 1	a. Visual cues b. Motion cues c. Aural cues d. Rudder and Nosewheel Steering e. Control stick f. Throttle/Engine response g. Attitude Indicator h. Performance Instruments	
Tech Order Climb (After attaining e[illegible]t airspeed and altitude, turn the shortest direction to intercept the 090 degree radial off the Bristol Lake VOR and climb to 10,000'MSL. FREEZE passing 9,000'MSL	RATING SCALE Use the following scale to compare the effectiveness of the simulator and the aircraft as training devices. The simulator would be a more effective training device than the aircraft. — YES → 5 NO ↓ The simulator would be just as effective as the aircraft. — YES → 4 NO ↓ The simulator would be less effective than the aircraft, but could be used for a major portion of the training. — YES → 3 NO ↓ The simulator would be less effective than the aircraft, but could be used for some portion of the training. — YES → 2 NO → 1	a. Visual cues b. Motion cues c. Aural cues d. Rudder e. Control stick f. Throttle/Engine response g. Attitude Indicator h. Performance Instruments i. Trim j. VOR/DME	

BEST AVAILABLE COPY

PAGE 1

C O P Y

MANEUVER DESCRIPTION	RATING SCALE	Circle any of the areas below in which differences existed between the simulator and the aircraft	COMMENTS
Proceed out the 090 degree radial at 10,000' MSL until entering a working area defined between 10-30 DME and the 080 to 100 degree radials.			
Normal Turns At 10,000', 160 KIAS, perform one 30 degree bank turn to the left, then one 30 degree bank turn to the right, then FREEZE the simulator	RATING SCALE Use the following scale to compare the effectiveness of the simulator and the aircraft as training devices. The simulator would be a more effective training device than the aircraft. — YES → 5 NO The simulator would be just as effective as the aircraft. — YES → 4 NO The simulator would be less effective than the aircraft, but could be used for a major portion of the training. — YES → 3 NO The simulator would be less effective than the aircraft, but could be used for some portion of the training. — YES → 2 NO → 1	a. Visual cues b. Motion cues c. Aural cues e. Control stick f. Throttle/Engine response g. Attitude Indicator h. Performance Instruments i. Trim	
Steep Turns At 10,000' 160 KIAS, perform one 60 degree bank turn to the left, then one 60 degree bank turn to the right, then FREEZE the simulator	RATING SCALE Use the following scale to compare the effectiveness of the simulator and the aircraft as training devices. The simulator would be a more effective training device than the aircraft. — YES → 5 NO The simulator would be just as effective as the aircraft. — YES → 4 NO The simulator would be less effective than the aircraft, but could be used for a major portion of the training. — YES → 3 NO The simulator would be less effective than the aircraft, but could be used for some portion of the training. — YES → 2 NO → 1	a. Visual cues b. Motion cues c. Aural cues e. Control stick f. Throttle/Engine response g. Attitude Indicator h. Performance Instruments i. Trim	

BEST AVAILABLE COPY

COPY

MANEUVER D. .RIPTION	RATING SCALE	Circle any of the areas below in which differences existed between the simulator and the aircraft	COMMEN
Constant Airspeed Climb 160 KIAS Climb to 12,000' ft, then FREEZE	**RATING SCALE** Use the following scale to compare the effectiveness of the simulator and the aircraft as training devices. The simulator would be a more effective training device than the aircraft. — YES → 5 NO The simulator would be just as effective as the aircraft. — YES → 4 NO The simulator would be less effective than the aircraft, but could be used for a major portion of the training. — YES → 3 NO The simulator would be less effective than the aircraft, but could be used for some portion of the training. — YES → 2 NO → 1	a. Visual cues b. Motion cues c. Aural cues e. Control stick f. Throttle/Engine response g. Attitude Indicator h. Performance Instruments i. Trim	
Constant Airspeed Descent-160 KIAS Descend to 10,000 ft, then FREEZE	**RATING SCALE** Use the following scale to compare the effectiveness of the simulator and the aircraft as training devices. The simulator would be a more effective training device than the aircraft. — YES → 5 NO The simulator would be just as effective as the aircraft. — YES → 4 NO The simulator would be less effective than the aircraft, but could be used for a major portion of the training. — YES → 3 NO The simulator would be less effective than the aircraft, but could be used for some portion of the training. — YES → 2 NO → 1	a. Visual cues b. Motion cues c. Aural cues e. Control stick f. Throttle/Engine response g. Attitude Indicator h. Performance Instruments i. Trim	

BEST AVAILABLE COPY

COPY

MANEUVER DESCRIPTION	RATING SCALE	Circle any of the areas below in which differences existed between the simulator and the aircraft	COMMENTS
<u>Airspeed increase 160 to 190 KIAS</u> Maintaining level flight at 10,000, increase airspeed from 160 to 190 KIAS, then FREEZE	<u>RATING SCALE</u> Use the following scale to compare the effectiveness of the simulator and the aircraft as training devices. The simulator would be a more effective training device than the aircraft. — YES → 5 NO The simulator would be just as effective as the aircraft. — YES → 4 NO The simulator would be less effective than the aircraft, but could be used for a major portion of the training. — YES → 3 NO The simulator would be less effective than the aircraft, but could be used for some portion of the training. — YES → 2 NO → 1	a. Visual cues b. Motion cues c. Aural cues e. Control stick f. Throttle/Engine response g. Attitude Indicator h. Performance Instruments i. Trim	
<u>Airpspeed decrease 190 to 160 KIAS</u> Level flight, 10,000 without speed brake, decrease the airspeed from 190 to 160 KIAS, then FREEZE	<u>RATING SCALE</u> Use the following scale to compare the effectiveness of the simulator and the aircraft as training devices. The simulator would be a more effective training device than the aircraft. — YES → 5 NO The simulator would be just as effective as the aircraft. — YES → 4 NO The simulator would be less effective than the aircraft, but could be used for a major portion of the training. — YES → 3 NO The simulator would be less effective than the aircraft, but could be used for some portion of the training. — YES → 2 NO → 1	a. Visual cues b. Motion cues c. Aural cues e. Control stick f. Throttle/Engine response g. Attitude Indicator h. Performance Instruments i. Trim	

BEST AVAILABLE COPY

COPY

MANEUVER DESCRIPTION	RATING SCALE	Circle any of the areas below in which differences existed between the simulator and the aircraft	COMMENTS
<u>Airspeed decrease from 190 to 160 KIAS</u> Level flight, 10,000 with speedbrake. FREEZE after trimming for 160.	**RATING SCALE** Use the following scale to compare the effectiveness of the simulator and the aircraft as training devices. The simulator would be a more effective training device than the aircraft. — YES → 5 NO ↓ The simulator would be just as effective as the aircraft. — YES → 4 NO ↓ The simulator would be less effective than the aircraft, but could be used for a major portion of the training. — YES → 3 NO ↓ The simulator would be less effective than the aircraft, but could be used for some portion of the training. — YES → 2 NO → 1	a. Visual cues b. Motion cues c. Aural cues e. Control stick f. Throttle/Engine response g. Attitude Indicator h. Performance Instruments I. Trim	
<u>Vertical S</u> Perform 1 ea vertical S A, and D, then FREEZE	**RATING SCALE** Use the following scale to compare the effectiveness of the simulator and the aircraft as training devices. The simulator would be a more effective training device than the aircraft. — YES → 5 NO ↓ The simulator would be just as effective as the aircraft. — YES → 4 NO ↓ The simulator would be less effective than the aircraft, but could be used for a major portion of the training. — YES → 3 NO ↓ The simulator would be less effective than the aircraft, but could be used for some portion of the training. — YES → 2 NO → 1	a. Visual cues b. Motion cues c. Aural cues e. Control stick f. Throttle/Engine response g. Attitude Indicator h. Performance Instruments i. Trim	

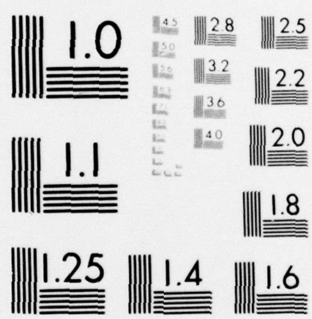
BEST AVAILABLE COPY

AD-A036 460 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OHIO SCH--ETC F/G 14/2
TECHNIQUES FOR THE INITIAL EVALUATION OF FLIGHT SIMULATOR EFFEC--ETC(U)
DEC 76 R L MILLER

UNCLASSIFIED GSM/SM/76D-34 NL

2 OF 2
ADA036460

END
DATE
FILMED
3 - 77



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

C O P Y

MANEUVER DESCRIPTION	RATING SCALE	Circle any of the areas below in which differences existed between the simulator and the aircraft	COMMENTS
Aileron Roll Perform an instrument aileron roll, then FREEZE	RATING SCALE Use the following scale to compare the effectiveness of the simulator and the aircraft as training devices. The simulator would be a more effective training device than the aircraft. — YES → 6 NO The simulator would be just as effective as the aircraft. — YES → 4 NO The simulator would be less effective than the aircraft, but could be used for a major portion of the training. — YES → 3 NO The simulator would be less effective than the aircraft, but could be used for some portion of the training. — YES → 2 NO → 1	b. Motion cues e. Control stick f. Throttle/Engine response g. Attitude Indicator h. Performance Instruments i. Trim	
Wingover Perform a wingover, then FREEZE	RATING SCALE Use the following scale to compare the effectiveness of the simulator and the aircraft as training devices. The simulator would be a more effective training device than the aircraft. — YES → 6 NO The simulator would be just as effective as the aircraft. — YES → 4 NO The simulator would be less effective than the aircraft, but could be used for a major portion of the training. — YES → 3 NO The simulator would be less effective than the aircraft, but could be used for some portion of the training. — YES → 2 NO → 1	b. Motion cues e. Control stick f. Throttle/Engine response g. Attitude Indicator h. Performance Instruments i. Trim	

PAGE 6

C O P Y

MANEUVER DESCRIPTION	RATING SCALE	Circle any of the areas *below in which differences* existed between the simulator and the aircraft	COMMENTS
Power-on Stall, Straight Ahead Enter and recover from a power-on stall, then FREEZE	RATING SCALE Use the following scale to compare the effectiveness of the simulator and the aircraft as training devices. The simulator would be a more effective training device than the aircraft. — YES → 5 NO The simulator would be just as effective as the aircraft. — YES → 4 NO The simulator would be less effective than the aircraft, but could be used for a major portion of the training. — YES → 3 NO The simulator would be less effective than the aircraft, but could be used for some portion of the training. — YES → 2 NO → 1	a. Visual cues b. Motion cues e. Control Stick f. Throttle/Engine response g. Attitude Indicator h. Performance Instruments i. Trim	
Power-on Stall, Turning Enter and recover from a power-on stall, then FREEZE	RATING SCALE Use the following scale to compare the effectiveness of the simulator and the aircraft as training devices. The simulator would be a more effective training device than the aircraft. — YES → 5 NO The simulator would be just as effective as the aircraft. — YES → 4 NO The simulator would be less effective than the aircraft, but could be used for a major portion of the training. — YES → 3 NO The simulator would be less effective than the aircraft, but could be used for some portion of the training. — YES → 2 NO → 1	a. Visual cues b. Motion cues e. Control Stick f. Throttle/Engine response g. Attitude Indicator h. Performance Instzuments i. Trim	

COPY

MANEUVER DESCRIPTION	RATING SCALE	Circle any of the areas below in which differences existed between the simulator and the aircraft	COMMENTS
Traffic Pattern Stall Series-Normal Begin the stall series at 15,000' MSL. FREEZE after completing the landing attitude stall	RATING SCALE Use the following scale to compare the effectiveness of the simulator and the aircraft as training devices. The simulator would be a more effective training device than the aircraft. — YES → 5 NO ↓ The simulator would be just as effective as the aircraft. — YES → 4 NO ↓ The simulator would be less effective than the aircraft, but could be used for a major portion of the training. — YES → 3 NO ↓ The simulator would be less effective than the aircraft, but could be used for some portion of the training. — YES → 2 NO → 1	a. Visual cues b. Motion cues e. Control stick f. Throttle/Engine response g. Attitude Indicator h. Performance Instruments i. Trim	
Slow Flight Fly for several moments in the 75-80 KIAS region and perform the following: a. Approach to a steep turn b. Abrupt control movement (pitch) c. Check control effectiveness d. Raising flaps to 50% note airspeed increase, then re-establish slow flight and fully retract flaps. Recover at first indication of stall, then FREEZE	RATING SCALE Use the following scale to compare the effectiveness of the simulator and the aircraft as training devices. The simulator would be a more effective training device than the aircraft. — YES → 5 NO ↓ The simulator would be just as effective as the aircraft. — YES → 4 NO ↓ The simulator would be less effective than the aircraft, but could be used for a major portion of the training. — YES → 3 NO ↓ The simulator would be less effective than the aircraft, but could be used for some portion of the training. — YES → 2 NO → 1	a. Visual cues b. Motion cues d. Rudder e. Control stick f. Throttle/Engine response g. Attitude Indicator h. Performance Instruments i. Trim	

PAGE 8

BEST AVAILABLE COPY

C O P Y

MANEUVER DESCRIPTION	RATING SCALE	Circle any of the areas below in which differences existed between the simulator and the aircraft	COMMENTS

Enroute Descent
Perform an enroute descent to 4,000 and proceed direct to the IAF for a VOR rwy 19 approach at Bristol Lake AFB (Approach shown on next page). FREEZE just prior to the IAF

RATING SCALE

Use the following scale to compare the effectiveness of the simulator and the aircraft as training devices.

- The simulator would be a more effective training device than the aircraft. — YES → 5
- NO ↓ The simulator would be just as effective as the aircraft. — YES → 4
- NO ↓ The simulator would be less effective than the aircraft, but could be used for a major portion of the training. — YES → 3
- NO ↓ The simulator would be less effective than the aircraft, but could be used for some portion of the training. — YES → 2
- NO → 1

a. Visual cues

b. Motion cues

c. Aural cues

e. Control stick

f. Throttle/Engine response

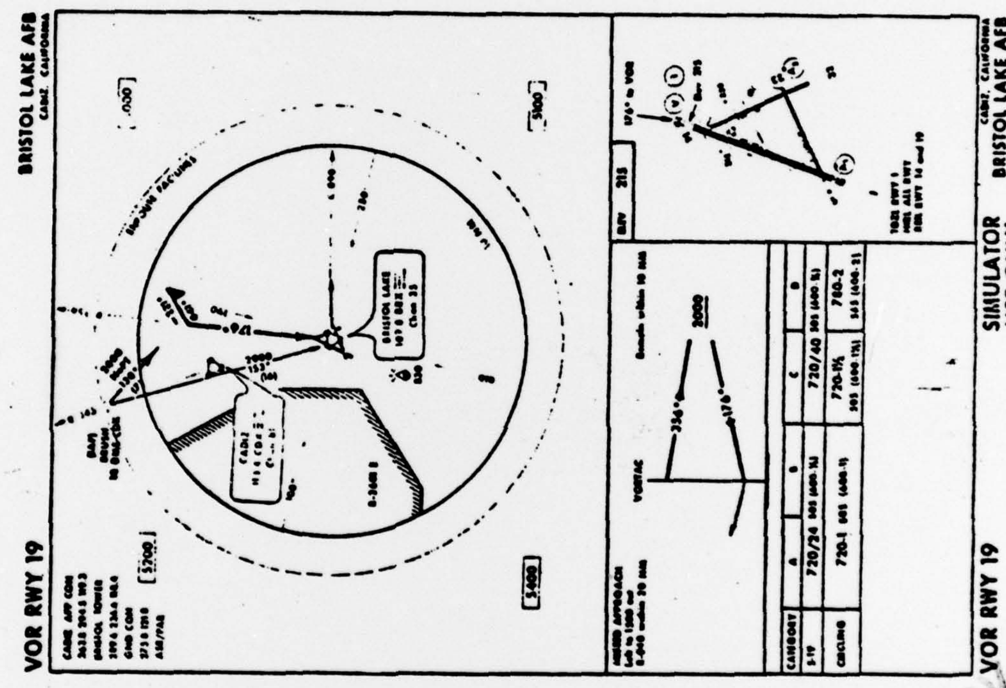
g. Attitude Indicator

h. Performance Instruments

i. Trim

BEST AVAILABLE COPY

C O P Y



10.

BEST AVAILABLE COPY

COPY

MANEUVER DESCRIPTION	RATING SCALE	Circle any of the areas below in which differences existed between the simulator and the aircraft	COMMENTS
VOR Approach Fly the VOR approach to a landing. Insure that you are down to minimum before the VOR which is the Missed Approach Point	RATING SCALE Use the following scale to compare the effectiveness of the simulator and the aircraft as training devices. The simulator would be a more effective training device than the aircraft. — YES → 6 NO ↓ The simulator would be just as effective as the aircraft. — YES → 4 NO ↓ The simulator would be less effective than the aircraft, but could be used for a major portion of the training. — YES → 3 NO ↓ The simulator would be less effective than the aircraft, but could be used for some portion of the training. — YES → 2 NO → 1	a. Visual cues b. Motion cues c. Aural cues d. Rudder e. Control Stick f. Throttle/Engine response g. Attitude Indicator h. Performance Instruments i. Trim j. VOR/DME	
Landing Land, select Motion Settle, open the canopy and deplane	RATING SCALE Use the following scale to compare the effectiveness of the simulator and the aircraft as training devices. The simulator would be a more effective training device than the aircraft. — YES → 6 NO ↓ The simulator would be just as effective as the aircraft. — YES → 4 NO ↓ The simulator would be less effective than the aircraft, but could be used for a major portion of the training. — YES → 3 NO ↓ The simulator would be less effective than the aircraft, but could be used for some portion of the training. — YES → 2 NO → 1	a. Visual cues b. Motion cues c. Aural cues d. Rudder e. Control stick f. Throttle/Engine response h. Performance Instruments i. Trim	

BEST AVAILABLE COPY

VITA

Ralph L. Miller was born on 22 November 1946 in Butler, Pennsylvania. He graduated from high school in Butler in 1964. He attended Grove City College in Grove City, Pennsylvania and graduated with a Bachelor of Science degree majoring in Electrical Engineering in 1968. He received a Reserve Officer Training Corp Commission in 1969 following ten months of employment as a plant engineer for the West Penn Power Company of Greensburg, Pennsylvania. He went on to Undergraduate Pilot Training at Williams Air Force Base, Arizona and received his wings in May of 1970. In the following 15 months he prepared for, and served, as a C130E crewmember in Southeast Asia. From September 1971 to June 1975, he served as a T-37 Instructor Pilot at Vance Air Force Base, Oklahoma (Air Training Command).

Permanent Address: R.D. #2
Butler, Pennsylvania

This thesis was typed by Miss Joyce L. Wilson.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER GSM/SM/76D-34	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER Master's thesis
4. TITLE (and Subtitle) TECHNIQUES FOR THE INITIAL EVALUATION OF FLIGHT SIMULATOR EFFECTIVENESS		5. TYPE OF REPORT & PERIOD COVERED M.S. Thesis
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Ralph L. Miller ~~Captain USAF~~		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Air Force Institute of Technology (AFIT-EN) Wright-Patterson AFB, Ohio 45433		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Institute of Technology (AFIT-EN) Wright-Patterson AFB, Ohio 45433		12. REPORT DATE Dec [redacted]76
		13. NUMBER OF PAGES 104 (104p.)
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

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

18. SUPPLEMENTARY NOTES

Approved for public release; IAW AFR 190-17.

JERRAL F. GUESS, Captain, USAF
Director of Information

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

Flight Simulator
Simulator Development
Simulator Evaluation
Transfer of Training
Fidelity of Simulation

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

This report presents an analysis of the development and initial evaluation of Air Force flight simulators. The objectives of the study were to determine the criterion variables most applicable to an initial flight simulator evaluation and to develop a general technique for the evaluation of these criterion variables.

The research began with a review of current Navy, Army, and Air Force flight simulator development and evaluation techniques. This review, combined with

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

UNCLASSIFIED 012 225

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) information gathered from related sources, provided the basis for examination and selection of criterion variables. The variables examined by this effort were: aircraft flight time saved, training efficiency, transfer of training, fidelity of psychological simulation, fidelity of engineering simulation, and simulator effectiveness. The examination of these variables concentrated on their measurability during an initial flight simulator evaluation and their ability to predict how well a flight simulator would perform its intended mission.

Following the examination of criterion variables, the research concentrated on the development of a technique for the evaluation of applicable criterion variables. The resulting technique is a combination of the traditional quantitative techniques plus some subjective techniques. The purpose of the subjective techniques is to identify simulator characteristics that are perceived to be different from the real work aircraft characteristics and to assess the impact that these differences will have on the operational use of the flight simulator.