To the Graduate Council:

I am submitting herewith a thesis written by Craig Shepperd Bowers entitled "Subjective Workload Assessment Techniques: A Test Pilot School Approach." I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Masters of Science, with a major in Aviation Systems.

R Richards

Mr. Robert Richards, Major Professor

We have read this thesis and recommend its acceptance:

Ralph D. Kimberlin

Dr. R. D. Kimberlin

C. T. Paludan

Dr. C. T. Paludan

Uwe Peter Solies

Dr. U. P. Solies

Accepted for the Council:

CW Minkel

Associate Vice Chancellor

and Dean of The Graduate School

SUBJECTIVE WORKLOAD ASSESSMENT TECHNIQUES: A TEST PILOT SCHOOL APPROACH

A Thesis Presented for the

Master of Science Degree

The University of Tennessee, Knoxville

Craig Shepperd Bowers

May, 1994

DEDICATION

To

Captain Steve Hazelrigg, USN,

whose legacy of leadership and loyalty

is surpassed only by the memory of his wonderful friendship.

ABSTRACT

This paper presents the results of research conducted to develop a program of instruction for subjective workload assessment techniques for test pilot students. The instruction program was developed to augment the current systems integration series in the Airborne Systems Department at the United States Naval Test Pilot School.

This paper introduces the need for more progressive instruction in workload assessment techniques during advanced cockpit systems integration and crew station evaluations. Material is presented to assist the test pilot school staff in developing a "Pilot Workload" course of instruction in order to introduce the student to the requirement for qualitative assessment techniques. A literature search was conducted to assess existing assessment techniques and analyze each for feasibility of incorporation into a test pilot school curriculum. A recommended program of instruction is included for academic and flight instruction which outlines the approach adopted by the United States Naval Test Pilot School as a means of increasing student training in subjective workload assessment techniques.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Advanced Systems Integration	ASI
Advisory Group for Aerospace Research and Development	AGARD
Bearing, Range, Altitude, Target aspect	BRAT
Circular Error Probability	CEP
Detailed Method Of Test	DMOT
Electronic Flight Information System	EFIS
Flight Mechanics Panel	FMP
Forward Looking Infrared	FLIR
Handling Quality Rating	HQR
Hands On Throttle And Stick	HOTAS
Head Up Display	HUD
Identification	ID
Instructor Pilot	IP
Instrument Landing System	ILS
Manned Flight Simulator	MFS
Measure of Effectiveness	MOE
Measure of Performance	MOP
Measure of Situational Awareness	MSA
Multi-Role Fighter	MRF
Nap Of the Earth	NOE
National Aeronautics and Space Administration	NASA
Naval Advanced Tactical Fighter	NATF
Naval Air Station	NAS
Naval Air Warfare Center	NAWC

North Atlantic Treaty Organization	NATO
Operational Evaluation	OPEVAL
Operational Test and Evaluation	OTE
Pilot-Vehicle Interface	PVI
Point Of Contact	POC
Quick Response Report	QRR
Situational Awareness	SA
Strike Aircraft Test Directorate	SATD
Subjective Workload Assessment Technique	SWAT
Task Load Index	TLX
Test and Evaluation Master Plan	TEMP
United States Air Force	USAF
United States Marine Corps	USMC
United States Naval Test Pilot School	USNTPS
United States Navy	USN
United States Ship	USS
Weapon System Operator	WSO

1. INTRODUCTION

1.1. ADVANCING TECHNOLOGY

Advances in communication, navigation, electromagnetic sensors, and weapons system technology have resulted in extremely complex cockpit designs with seemingly unlimited potential. Current flight information management systems, radar systems, and tactical weapons devices are becoming increasingly complex and more automated each year. Today, the limiting factor in man/machine interface is more often man's capacity rather than machine potential. Possibly the most obvious example of this is the single seat multi-role tactical aircraft cockpit: the human operator is quickly becoming overloaded with information that new technology is providing.

Technological advances will continue to push aircraft designers, as military and commercial spending continues to decline after the boom period of the last decade. Consider programs eliminated in large part due to budgetary considerations in the last several years: the Navy A-12, P-7, and Naval Advanced Tactical Fighter (NATF) and the Air Force Multi-Role Fighter (MRF). While there are examples of new high visibility acquisition programs today (V-22, C-17, and F-22), even these are facing fiscal hurdles in the wake of lower projected buys and increased unit costs. One result of lower defense spending will be an increase in system improvement programs, or upgrades. Currently, numerous programs are being considered to help bridge the gap between aging fleet aircraft and limited procurement dollars: F-14, E-2, F-16, and the B-737 to name a few.

The potential for reduced test and evaluation schedules and budgets, coupled with the ever-increasing advances in design technology dictate that evaluators must spend precious dollars and time wisely. Evaluation methodology must be in place throughout the design or improvement process to ensure that money being spent for new hardware to the fleet will be of some benefit upon its arrival.

1.2. CREW STATION EVALUATION

As modern technology provides more capable aircraft sensors and weapons systems, evaluators must keep pace with their ability to provide effective feedback throughout the design process. In this case, the term evaluator encompasses the entire spectrum of airborne system operators: pilots, weapon system officers, engineers, navigators, etc. Crew station evaluation can encompass many varied areas: system performance, system integration, and Pilot-Vehicle Interface (PVI). System performance using established quantitative test techniques receives a lot of emphasis in most flight test instruction programs. Human factor issues such as system integration and PVI are much more difficult to quantify and do not receive as much focus. Obviously, subjective opinion is more difficult to support than quantitative test data - many times there are no strip charts or data plots to support aircrew opinion. In the late 1960's, R. P. Cooper and G. E. Harper published the Cooper-Harper Handling Qualities Rating Scale as a method of quantifying pilot opinion of aerospace vehicle handling qualities. The HQR Scale, as it is widely known, has been accepted throughout the test and evaluation community, providing a common reference frame for use between aircrew and engineers in discussing aircraft flying qualities.

In systems test and evaluation, however, it is difficult to draw an analogy to the Cooper-Harper HQR Scale in terms of a technique that has earned such widespread acceptance and utilization. While there are several techniques in existence that appear to hold great promise, none has become the benchmark in the flight test instruction community that the HQR Scale is in flying qualities training. Specifically, the United States Naval Test Pilot School (USNTPS) is attempting to increase student instruction in subjective assessment techniques in hopes of better preparing aircrew for the increasing systems integration evaluation workload expected in military test and evaluation in the near future. The following techniques and instructional program will provide a more advanced academic and flight instruction syllabus for subjective workload assessment in advanced weapons systems integration and crew station evaluations.

2. BACKGROUND

2.1. UNITED STATES NAVAL TEST PILOT SCHOOL

Located at Naval Air Station Patuxent River, MD, USNTPS serves to advance the specialized instruction in aircraft weapons systems test and evaluation and acquisition support. The school consists of three curricula (Fixed Wing, Rotary Wing, and Airborne Systems) which run concurrently over an eleven month period involving numerous academic and flight exercises which introduce the student to the concepts involved with flight test theory and evaluation techniques. Two classes are conducted each year, with each class containing about 35 students comprised of United States Government and Foreign Military Service personnel.

2.1.1. Airborne Systems Training

The USNTPS Airborne Systems Department is charged with instructing advanced airborne systems test and evaluation flight test techniques including communication, navigation, radar, and electromagnetic sensors. The curriculum is designed to provide future Project Pilots, Naval Flight Officers and Flight Test Engineers instruction in current and future aircraft and weapon systems in preparation for placement at any one of several test directorates within the Naval Air Warfare Center (NAWC) organization, including the Aircraft Division based at NAS Patuxent River, MD and the Weapons Division based at NAWC China Lake, CA. Close liaison with test activities and a steady rotation of second-tour instructor test pilots enables USNTPS to incorporate new flight exercises as required to provide adequately trained aircrew and engineers to field test facilities. Upon completion of the course, the graduate will have demonstrated the ability to satisfactorily develop, execute, and report on a comprehensive flight test program on a state-of-the-art aircraft platform simulating a Navy Developmental Test - Phase IIA Evaluation requiring a report on aircraft systems, performance, and flying qualities.

Only Naval Flight Officers and Flight Test Engineers are enrolled in the Airborne Systems syllabus; pilots are placed in the either the Fixed-Wing or Rotary-Wing syllabus depending on their fleet experience. The ability to qualitatively evaluate an airborne weapon system or advanced cockpit design is not limited to any particular aircraft seat: pilots, navigators, and weapon system officers should all be well-versed on current techniques. However, at USNTPS, because of the segmented approach to systems evaluation, the Fixed-Wing syllabus deserves a closer look as it represents many of the limits to effective training in this area. Thus, the Fixed-Wing syllabus served as the catalyst for this research though the fundamentals are applicable to all airborne systems training.

During the 11-month course, the student test pilot receives over 500 hours of academic instruction and 79 flights which include only three exercises dedicated to Airborne Systems evaluation. Unfortunately, the heavy course load of the Fixed-Wing curriculum does not allow for students to break off for several weeks of system training; all systems work must be integrated into the Fixed-Wing schedule. The result is a segmented approach to systems evaluation techniques that often experiences several months lapse between successive flight exercises. In an attempt to keep a building-block approach for student test pilots, the Advanced System Integration (ASI) series was established consisting of three flight exercises:

> **ASI I** - air-to-ground radar evaluation with emphasis on controls, displays (Head-Up (HUD) and Multi-Function), and cockpit human factors,
>
> **ASI II** - air-to-air radar evaluation with emphasis on project teamwork and mission suitability, and
>
> **ASI III** - night vision device demonstration with emphasis on compatibility issues.

Within the constraints of aircraft availability and budget limitations, these exercises have been designed to develop the student's ability to approach systems testing and conduct flight testing as required to validate system performance. Content of academic instruction is monitored to provide supporting system operation and test technique theory as applicable. However, upon review of learning objectives associated with the ASI series, it was evident that quantitative test techniques dominated student instruction and very little training was being conducted in qualitative assessment techniques. A closer look at student instructional material revealed that all Airborne Systems training (Pilot, Naval Flight Officer, and Flight Test Engineer) could benefit from exposure to qualitative assessment methods. The advantage of this approach is that it builds on handling qualities evaluation techniques and addresses the particular requirements of airborne systems evaluations.

2.1.2. Emphasis on Quantitative Testing

Historically, student training in the Airborne Systems syllabus has centered on quantitative test techniques as a means of evaluating system performance. The primary training document for systems testing at USNTPS is the Systems Test Manual (Preliminary), which was originally published in 1992. The manual is well-written and gives the student a cookbook approach to system flight test techniques, providing a by-the-numbers guide to the entire process from test theory to preflight planning, flight procedures, and data reduction. However, the manual reflects a strong emphasis on quantitative testing with little guidance in qualitative test techniques. For instance, Section III lists the following air-to-ground and air-to-air radar test techniques (Landmannn, 1992):

- Preflight and Built-In Tests
- Antenna Stabilization
- Antenna Scan Rates
- Antenna Scan Angle Limits
- Elevation Angle Limits
- Minimum Range
- Range/Azimuth Resolution
- Velocity Resolution (A)
- Maximum Detection Range
- Probability of Detection (A)
- Range/Bearing Accuracy
- Doppler Beam Sharpened Notch Width
- Antenna Tracking Rate Limits (A)

Blind Velocities/Ranges (A)	(A) Air-to-Air
(G) Air-to-Ground	

Qualitative techniques, in comparison, are limited to just three: Mapping Quality (G), System Integration, and Mission Utility. The focus on quantitative systems testing is repeated in Section V: "Electro Optical Systems - FLIR Evaluation". Consider the following list of quantitative techniques presented:

Preflight and Cooldown	Line Of Sight Stability
Controls/Display	Airspeed effect
Boresight Accuracy	Stabilization Limits
Slew Limits/Rates	Detection/Classification/Identification
Field Of View	Minimum Tracking Range
Line Of Sight Drift Rate	
Static/Dynamic Resolution	

Only two qualitative issues, System Integration and Mission Utility, are presented to the reader. Not surprisingly, the results of student test plans and reports is a heavy slant towards the quantitative performance of airborne systems with no effective means of evaluating mission utility considerations.

While quantitative test techniques are an important part of the evaluator's skills, the ability to adequately test for specification compliance (system performance) should not overshadow the importance of qualitative evaluation techniques. This is where the importance of qualitatively assessing mission suitability through crew station design, system integration, and mission utility assessment techniques cannot be overstated. Consider the flying qualities approach: while the stick force per g of a potential trainer aircraft may meet specification threshold, a test pilot whose arm is unduly fatigued after conducting a loop maneuver will probably frown on the aircraft's suitability for extended acrobatic maneuvers in the student syllabus. Likewise, an air-to-air radar that can routinely detect a section of enemy fighter aircraft at 100 miles may sound like a tremendous asset, but it is useless if the operator is unable to control cursor positioning or select the desired operation mode. To be sure, quantitative data is important as a means of ensuring specification requirements have been met, but it is mission suitability that must be satisfactory if the system is going to pass Operational Evaluation (OPEVAL) milestones in the Naval Aviation acquisition process.

While listing the techniques included in student instruction, it is also necessary to comment on the content of that training. Qualitative evaluation techniques are not discussed in great detail in the System Test Manual other than "using comments/remarks section of individual tests to include qualitative comments about integration problems or contributions." Section V, "Electro-Optic/FLIR Evaluation", lists the procedure for System Integration/Mission Utility as "Fly mission representative profiles to exercise the Forward Looking Infrared (FLIR) integration with the radar, navigation, and other weapon system components as appropriate." The limited focus in qualitative techniques carries over to flight exercise learning objectives which contain the same emphasis on quantitative testing, providing the student with little guidance on how to support subjective opinion. Maximum detection range, scan angle limits, and minimum resolvable temperature differential of a FLIR system can be quantified and as such are easy to instruct and grade student technique. Subjective assessment, or pilot opinion, is a completely different story. Subjective assessments are difficult to quantify and more difficult to instruct how to obtain one. There is no way to objectively evaluate a student's opinion, other than addressing quantitative test results used in support of their position. Furthermore, a test pilot needs some technique for quantifying the comment like "good", "effective", or "difficult" to objectively relay problems associated with system design to become an effective contribution to the flight test program.

2.1.3. Supporting Pilot Opinion

As technological advances ensure further developments in increased tactical capability through improved avionics systems and cockpit design, it is important that the human operator be able to effectively operate the end product. In fact, an argument can be made that system integration evaluation should be at the beginning of design review, as system integration issues that are not solved in the beginning may become unworkable later in the production schedule. End product effectiveness dictates that system integration problems should be corrected early in the process: the new radar system in the AV-8 aircraft will be useless if the control and display design and integration are so poor that the pilot cannot operate the system well enough to detect enemy aircraft. The problem is by no means endemic to military aviation alone as evidenced by the focus given to avionics design in the new Boeing transport, the 777. Components such as the Electronic Flight Information System (EFIS) have received extensive design evaluation and review as reported in *Aviation Week and Space Technology* (O'Lone, 1992).

USNTPS is not the only institution attempting to address the shortcoming in instructional technique for subjective assessment techniques in aircraft design. In fact, discussions with instructors at the USAF Test Pilot School at Edwards Air Force Base, CA and the Empire Test Pilot School in Boscombe Down, England indicated a similar desire for improving the approach to subjective assessment technique instruction (Abraham, 1994, and Spear, 1994, respectively).

Several techniques exist which allow aircrew to relate difficulties experienced in employing aircraft or systems operationally which could be integrated into the test pilot curriculum to fill the void of subjective assessment techniques. The author's experience while instructing in the USNTPS Airborne systems course indicates that many students can relate an inability to employ an aircraft control or weapon system to a problem of task loading: there just wasn't enough time to analyze the situation, decide on a course of action, and complete it before aircraft or weapon limitations were exceeded. In short, if the student could quantify the workload of various flight tasks, supporting pilot subjective opinion with quantifiable data would assist in relating system design deficiencies. Pilot workload assessment can be applied over a wide range of flight tasks and can be used to address a variety of design issues like crew station design, system integration, and mission utility. If the pilot could not maneuver the aircraft to a firing position because readying the weapon for launch required all available factions, the system won't work. In this case, discussing problems in terms of pilot workload would clearly assist the test pilot in discussing design problems.

To that end, research was conducted to evaluate existing techniques for subjective workload assessment and analyzed for incorporation into the USNTPS curriculum. Techniques needed to be well documented, easy to use, and readily adaptable to classroom instruction and flight exercises. The result was the techniques introduced in Section Four and a recommended Program of Instruction as presented in Section Five.

3. EVALUATION METHODOLOGY

3.1. PILOT IN THE LOOP

Aircrew opinion is held in high regard in the development of military aircraft. Operational Test and Evaluation (OTE) is a key step in most military aviation Test and Evaluation Master Plans (TEMP) and a lot of importance is placed on successful completion of the OTE milestone. The trend today is to get the aircrew (pilot, navigator, weapon system officer, etc.) involved in crew station analysis as early as possible in the design phase to prevent design flaws from becoming so firmly embedded in the manufacturing process that corrections downstream become too costly to correct. Aircrew Systems Advisory Panels and Design Review Boards are just two examples of attempts to get aircrew involved early in the development stage. The tactical importance and financial benefit of allowing the operator to play an active role in the development process through ongoing evaluation of the system cannot be overstated.

Simulator use in the design review phase has increased in the recent past; this trend is likely to continue for two reasons which are both directly related to cost-reduction: early warning of potential problems and lower cost per flight hour than airborne testing. Flight test crews must be able to effectively incorporate simulators into flight test plans not only as a means of reducing project flight hour costs but also to gain insight into potential design deficiencies early in the design process.

3.2. SIMULATION

The United States Air Force has devoted a great deal of effort in developing effective crew station design evaluation techniques. The USAF Handbook for Conducting Pilot In the Loop Simulation for Crew Station Evaluation contains a great deal of information for test crews to consider when tasked with using pilot-in-the-loop flight simulation as a means to assess system integration effectiveness (Lehman et al, 1989). The USAF approach includes some fundamental

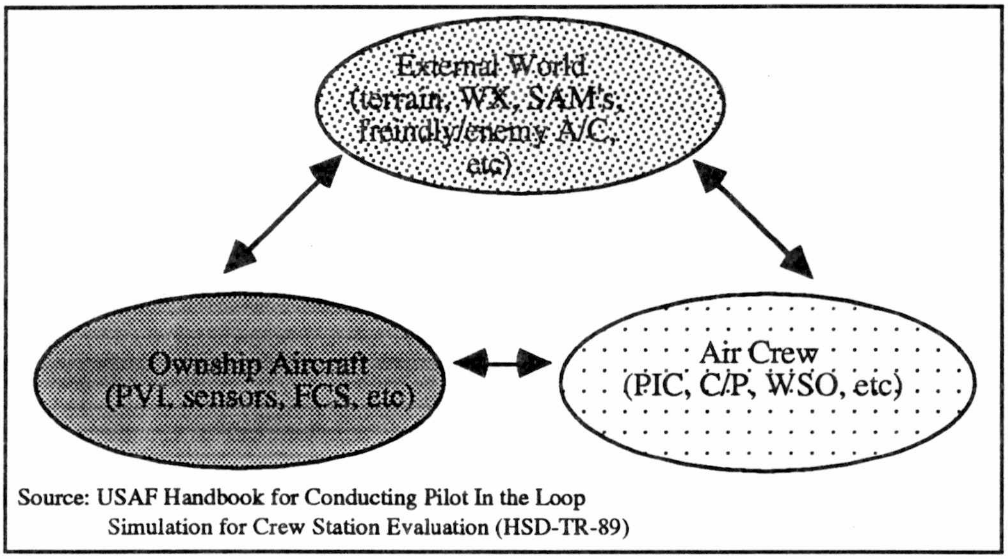


Figure 3-1

Evaluation Environmental Factors

concepts that should be included in any systems evaluation brief for student evaluators. Figure 3-1 helps to depict the interaction between pilot and environment factors which must be considered in any crew station design evaluation. It should be noted that while specific examples contained herein refer to piloting tasks, "pilot" can be replaced with "system operator" as the argument applies to anyone tasked with operating an airborne system.

The first priority in any test program must be clearly defined test objectives and student evaluator instruction should emphasize this. In a test pilot school syllabus exercise, the student evaluator should be required to develop the flight test objectives very early in the test planning process. In fact, asking the test team to formulate a list of mission suitability requirements and then develop test objectives will ensure that the test team remains focused on testing to the performance capabilities that it has established. The test objectives should be specific enough so as to be measurable, both in terms of system performance and Pilot/Vehicle Interface, or PVI. The PVI becomes a limiting factor if, for instance, the radar is <u>capable</u> of detecting a specific target at 75 miles but a WSO is unable to detect any target outside of 40 miles due to poor cursor controller sensitivity and display resolution. One is purely a function of radar system performance, the other is a function of poor PVI.

Test aircrew should be cautioned against ambiguous test objectives such as "qualitative assessment of system integration" as this provides very little objective requirements by which to rate the system design. With some focused discussion, a group of experienced aviators could certainly develop quantifiable standards that should be met to qualify as "good system integration." A test objective such as "ability to track three target aircraft within 40 nautical miles and deliver a target position report within five seconds of acquisition for each aircraft" is one example of an attempt to qualify what is acceptable performance of the man/machine system. In that sense, multi-tasking can be a very effective way of evaluating design effectiveness. Communication tasking or navigation updates during target prosecution are examples of multi-tasking requirements that can aid in evaluating system integration. Understandably, evaluating an advanced tactical aircraft crew station is going to require an extensive test matrix with numerous complex tasks in order to build a mission representative scenario. It is important that the student evaluator gain an appreciation for clear test objectives that can be quantified in order to provide some standard of measurement for system effectiveness.

Once the evaluation objectives are identified, the determination must be made as to which techniques best suit the data requirement. The relationship between performance, workload, and Situational Awareness (SA) evaluation techniques as presented in reference (5) is an important one. Figure A-1 depicts how subjective data fits into the equation of "mission effectiveness" which is identified as a function of more limited areas of quantitative evaluation including performance and workload assessments. A review of Measure of Effectiveness and its composite parts will help to impress student evaluators as to the importance of workload measurement techniques.

3.2.1. Measure of Effectiveness

A Measure of Effectiveness (MOE) could be considered the indicator of PVI operational utility and is a combination of performance, workload, and situational awareness. MOE's address the extent to which an item contributes to accomplishing the assigned mission or task, and may be quantitative (aircraft loss rates in simulated engagements, Bomb Damage Assessment via No Drop Bomb Scoring System, etc.) or qualitative (aircrew or expert observer subjective assessment). The subject of the evaluation could be an aircraft system (entire air-to-air weapon system), a subsystem (fire control radar), or device (digital display). For the USNTPS ASI II exercise, system effectiveness in the air-to-air mission is the goal with workload assessment being the focus of the evaluation. However, quantitative test techniques as a measure of performance are important tools and will be utilized as well. Situational Awareness evaluation techniques are mentioned here though are out of scope of the ASI series exercises.

3.2.2. Measure of Performance

A Measure of Performance (MOP) is an indicator of the behavior or capability of a system, subsystem, pilot or device consisting primarily of quantitative measures relating to system (detection/tracking range) and pilot (reaction time, frequency of occurrence) performance. MOPs are subdivided into 5 categories:

Weapons - CEP, Miss Distance, Kills/Missile Fired

Sensors - Target ID Range, Maximum Detection Range, False Alarm Rate,

Aircraft Control - Deviation from intended glideslope, airspeed, or navigation waypoint. Performance rating based on ability to obtain positional advantage.

PVI - Effectiveness measures related to pilot performance and mission effectiveness. Objective pilot performance techniques may be divided into two categories:

System plus operator: gun-sight tracking error, engine thrust, control surface deflection,.

Direct: response times, communication delay, control movements per second.

Subjective - consist of pilot or expert observer questionnaire/interview/rating scales

3.2.3. Measure of Situational Awareness

A Measure of Situational Awareness (MSA) is a technique which seeks to quantify aircrew comprehension of the surrounding environment: in fact, what reality is as the aircrew perceives it. Currently, MSA techniques are undergoing extensive research due to their multi-dimensional nature and much discussion over exactly what constitutes "good" SA. As a result, MSA techniques are not within the scope of the ASI series at USNTPS. The USAF Handbook provides insight into subjective and objective measurement techniques and evaluators should be aware that there is a need to consider aircrew SA in a comprehensive evaluation of system effectiveness.

3.2.4. Measures of Workload

Aircrew workload can be considered the combined effect of system performance, situational awareness, and additional effort associated with any particular mission task. A Measure of Workload (MOW) attempts to identify aircrew mental or physical tasking and serves as an excellent measure of the synergistic effect of crew station design and system integration. A more detailed discussion of the concept of aircrew workload and its effect on mission utility are warranted here.

3.3. AIRCREW WORKLOAD

Workload has been a chief concern to aircraft designers for many years. In fact, as long as there is a human in the system operation, workload is going to be of interest to the design engineer. The Cooper-Harper HQR Scale reflects the importance of pilot workload in handling qualities evaluation by considering pilot workload in the very first question once an aircraft is considered controllable with the question: "Is adequate performance attainable with a tolerable pilot workload?" (Cooper, 1969)

The Advisory Group for Aerospace Research & Development (AGARD) was formed to bring together leading NATO experts in the field of science and technology relating to aerospace for the purposes, among others, of exchanging information, stimulating scientific advances, and improving aerospace research and development cooperation. In 1980 a proposal requesting further study of pilot workload was forwarded to the AGARD Flight Mechanics Panel (FMP). The proposal highlighted that despite extensive effort devoted to the subject, little progress had been made towards "formulating a readily acceptable definition of the term pilot workload nor towards recognizing suitable techniques for assessing levels of workload." (Woodfield and Renaudie, 1987) The proposal recommended three specific tasks which could be considered in the evaluation of pilot workload:

Helicopter Nap of the Earth (NOE) Flight Maneuvers

Tactical Weapons Delivery Maneuver in Ground Attack Aircraft

Jet Transport Manually Flown ILS Approach

Appendix B lists the tasks involved in the tactical weapons delivery profile to give an example of maneuvers required in a workload assessment scenario.

The FMP accepted the proposal in May 1981 and invited anyone interested in the subject to participate in the study. The results of the study were published in July 1987 in AGARD Report No. AGARD-AG-282, reference (6), which reported that while the term pilot workload is being used more and more by designers and operators of modern aircraft, there is still no generally accepted definition of the term nor is there agreement on how best to assess it. In an attempt to fill this void, a modified version of pilot workload definition was proposed: "Pilot workload is the integrated mental and physical effort required to satisfy the perceived demands of a specified flight task," (Ellis and Roscoe, 1985).

3.4. WORKLOAD MANAGEMENT

Workload level must be managed throughout the design stage, whether in a system upgrade or new design process. The requirement to refrain workload from degrading aircrew effectiveness is no less important in the radar/avionics upgrade in the F-14 as it is in the cockpit design of the F-22. In either case, designing a system that cannot be employed by the fleet operator would seriously limit its chances for successful completion of the Operational Evaluation (OPEVAL) and render the entire acquisition program useless. The importance of considering pilot workload as early as possible in the design phase was highlighted by D.W.Milam (1982): "Workload measurements, whether subjective or objective, should be available much earlier in the design process so that design options can be intelligently considered."

The idea of workload management should not be considered as solely a tactical aircraft design problem, either. Technological advances in civil transport autopilot systems, flight management computers, and cockpit displays has introduced new demands on flight deck personnel that must be optimized, if not reduced, during critical phases of flight. Certain cases have pointed to potentially tragic consequences if the pilot workload issue is not carefully managed, as was indicated in the crash of a French A320 crash near Strasbourg, France in 1992. The investigation identified complex human factor issues that contributed to the accident, "underscoring the need to more fully understand the implications of man/machine interface as increasingly advanced technologies are used on civil transport aircraft." (Sparaco, 1987) Boeing Aircraft Company has focused on flight deck design problems in the development of its new 777 transport aircraft, also. By keeping the pilot in the loop during design phase and evaluating workload concerns, the engineers have been able to solve several problems simultaneously. One instance was with electronic checklists, which are displayed on 8x8 inch liquid crystal flat panel color displays. Through the use of PIL evaluations, engineers determined that the system could remove a lot of the workload from the pilot, yet keep them in the management loop by requiring action on their part before the display changes. By keeping the pilot in the loop yet removing pilot workload, engineers have benefited on two fronts through the subjective assessment of pilot workload.

3.5. WORKLOAD ASSESSMENT

Numerous techniques exist for assessing pilot workload. The need to assess workload in combat aircraft during flight can pose special problems, as some techniques which may be suitable for the relatively "safe" profile flown by a civil transport does not suit the dynamic flight environment of a tactical fighter aircraft. In fact, several criteria for workload assessment techniques have been generally accepted, including safety, operator acceptance, and convenience, reliability, and intrusiveness. The workload assessment techniques can be loosely classified into three groups: objective, subjective, and physiological, .

3.5.1. Objective Assessment

Objective assessment techniques, those not influenced by personal feelings, can be divided into two broad categories: performance and analytical.

3.5.1.1. Performance

The relationship between man/machine performance and pilot workload is not a simple one, yet very real. If a pilot is extremely occupied with various tasks, chances are glideslope tracking performance and desired runway point of interception are going to suffer. Likewise, a pilot that needs to continually monitor radar operational status and Built-In Test indicators is likely to lose valuable tracking time due to delayed initial acquisition of threat aircraft. Not that this will happen every engagement, but probability is high that over time, man/machine performance will suffer under the increased workload. Performance measures attempt to quantify those errors that are indisputable; landing 200 feet long, 2 degrees above glideslope for 38% of the approach, and 90% probability of max detection decreasing from 38 miles to 31 miles. The opposite could also be true, as a pilot's performance may actually decrease in low-workload states due to complacency.

A number of techniques are available ranging from extremely sophisticated measuring devices (instrumented aircraft, external measuring devices, theodolite) to extremely simple (stopwatch, tape measure, video camera). Another performance measurement technique is a secondary task technique, which evaluates a pilot's performance of additional tasking (i.e. - radio communications) while the pilot is concentrating on the primary task (i.e. - glideslope tracking). More information can be obtained on these techniques in the USAF Handbook for Conducting Pilot-In-The-Loop Simulation for Crew Station Evaluation.

3.5.1.2. Analytical

One interpretation of workload is that it increases with the time demands of a specific task. In this sense, a task that requires 38 seconds to complete must increase the workload of the pilot more than a task which takes 21 seconds. Using this definition, analytical techniques consist primarily of time-motion studies of cockpit tasks to determine what tasks are higher workload than others. This technique finds increased application in design considerations in crew-station layout and aircraft certification procedures.

3.5.2. Subjective Assessment Techniques

The importance of aircrew subjective assessment can be summed up very easily: "If an operator *feels* his workload level is high, then it *is* high, regardless of what other measures show. " (Johannsen et al, 1979). This technique is perhaps the most beneficial means of evaluating pilot workload in existence. As Roscoe pointed out, "Subjective reporting, by experienced test pilots is undoubtedly the most commonly used and probably the most reliable way of assessing workload in flight presently available. . . subjective impressions of how hard a pilot is working - the amount of effort he has to exert to meet the demands of the task - are most relevant." (Roscoe, 1987) The two most common subjective assessment techniques are rating scales and post flight interviews.

3.5.2.1. Rating Scales

The most well known rating scale, the Cooper-Harper, has often been mistakenly used for workload evaluations. This is with good reason as the advantages are numerous: pilots need little training in its application, it is designed for in-flight use, and it is safe. However, the Cooper-Harper was designed as a flying qualities evaluation tool and as such is geared towards psychomotor tasks and ratings based on pilot compensation to achieve desired or adequate aircraft performance. In fact, several rating scales have been developed which are better suited for workload assessment. Introducing student test pilots, flight officers, and engineers to these other techniques is necessary to ensure that future Project Officers and Project Engineers use the best tools available in system integration evaluation programs. The techniques discussed in this paper focus on rating scale formats including the Modified Cooper-Harper, Honeywell Cooper-Harper, and Bedford Rating Scales and two multi-dimensional techniques, the Subjective Workload Assessment Technique and NASA Task Load Index.

3.5.2.2. Interviews

Post-flight interviews or questionnaires are another common form of subjective workload analysis. An interview conducted immediately after the flight can be extremely quick and easy to conduct, and are highly effective. Questions should be developed which attempt to solicit aircrew comments that might otherwise go unnoticed during the course of a rating scale evaluation. It may take some pointed questioning to draw the real root of task overload from an aviator's memory bank after the flight is completed. Questions may be open- or closed-ended and need to be carefully worded to ensure the responses will be relevant to the desired mission tasking under consideration. The post-flight interview can act as a good complement to in-flight techniques: the rating scale does not rely heavily on aircrew memory yet provides limited diagnostic capability, the interview provides the test team with detailed aircrew analysis and supporting data for flight test ratings.

During the USNTPS Airborne Systems Integration II exercise, student test pilots are required to develop a five item questionnaire to be completed during the flight debrief. Questions are evaluated by the staff for applicability and usefulness in supporting in-flight results.

3.5.3. Physiological Assessment Techniques

Physiological techniques include monitoring pilot heartbeat, brain function electrical activity, and physical effort have employed to assess aircrew workload levels and are getting increasing use as a means of supporting subjective assessment data. These techniques are beyond the scope of this paper, though much information is available in numerous reports and manuals (Lehman et al, 1989).

3.6. SUMMARY

Pilot in the loop evaluation methods will be an increasing part of system evaluation. Test crews should be fluent in subjective workload assessment techniques available to them. Rating scales and interviews are two commonly used techniques that deserve mention in any test pilot curriculum. Several advantages to subjective testing include non-intrusiveness, ease of implementation, operator acceptance, and easily administered. Numerous disadvantages have been reported including a limited capability for statistical analysis and task-specific results that can not be applied to other flight maneuvers. O'Donnell and Eggemeier observed that reliance on short term memory and tendency for evaluator to dissociate subjective rating from task performance were also limits of subjective assessment techniques (O'Donnell and Eggemeier, 1986). While these limits do not significantly affect a test pilot school flight exercise, they are all valid limitations and should be discussed with the student evaluators.

4. TECHNICAL REVIEW

4.1 HANDLING QUALITIES MODEL

Subjective assessments of aircraft handling qualities will be of concern to engineers as long as there is human-controlled flight. As important as operator comments are, a problem arises when one pilot's opinion is compared to another. Without some standard of measurement, there is no way to normalize subjective evaluations in order to give design engineers a consistent, credible source of feedback. Since the early 1960's, a lot of attention has been given to developing a standardized scale of measurement for subjective assessment of aircraft handling qualities in order to assign a qualitative number to qualitative data. In 1969, George Cooper of NASA Ames Research Center and Roger Harper of Cornell Aeronautical Laboratory published the Cooper-Harper Handling Qualities Rating (HQR) Scale with the purpose to "re-examine existing rating scales with the purpose of developing a single improved scale and of clarifying its use in the evaluation of handling qualities." (Cooper and Harper, 1969) Since its introduction, the Cooper-Harper HQR Scale has become the foundation for subjective assessment of aircraft handling qualities. Handling quality ratings are routinely used in ready room discussions, during test flights, post-flight analysis, and in technical reports. While there is much argument regarding the applicability of the Cooper-Harper scale in areas other than handling qualities, most subjective assessment techniques are based in part on the design of the Cooper-Harper HQR Scale. It is important, therefore, that any discussion of subjective workload assessment techniques begin with a review of Cooper-Harper fundamentals.

4.2. COOPER-HARPER HANDLING QUALITIES RATING SCALE

4.2.1. Background

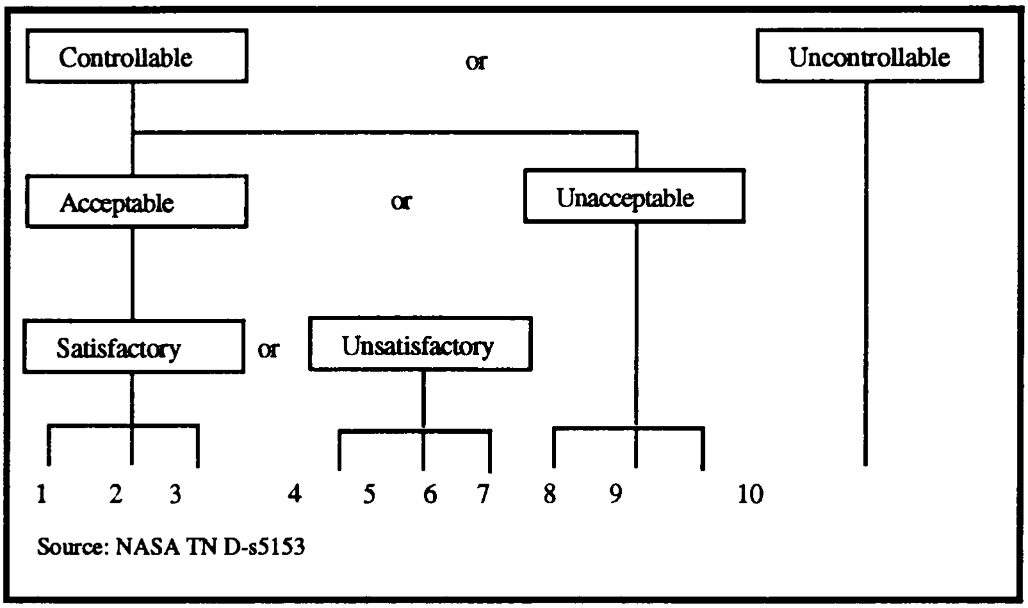
The Cooper-Harper Handling Qualities Rating (HQR) Scale depicted in figure A-2 was first published in NASA Technical Note D-5153 in April of 1969. As Cooper and Harper stated, for a pilot rating scale to be universally acceptable and consistently applied in the evaluation of handling qualities the terminology must be easily understood by everyone in the field. Several key terms were defined and are summarized here.

> Handling Qualities are those qualities or characteristics of an aircraft that govern the ease and precision with which a pilot is able to perform the tasks required in support of an aircraft role. Factors which influence handling qualities are stability and control characteristics, environment, cockpit interface, and stress.
>
> Mission is a general term used to convey the concept of operational requirements, and includes all functions that the pilot-vehicle combination must be able to accomplish. Mission should not be confused with task.
>
> Task is the actual work that the pilot is assigned in completion of a designated flight segment to include control tasks (actuation of principal controls) and auxiliary tasks (pilot actions other than in direct control of the aircraft.)
>
> Flight Phase is used to delineate a particular segment of flight for more definitive examination, such as Ground/Deck, Cruise, and Approach.
>
> Sub Phase is the part of a flight phase which has a single objective, such as Confined Area Positioning in the Ground/Deck Phase.

Aircraft mission suitability may be examined by theoretical analysis, experimental performance measurement, and pilot evaluation. Cooper and Harper were quick to note that while satisfactory mission performance is a fundamental reason for the concern with handling qualities, experimental measurement of pilot-vehicle output performance is not always the end all to assessing aircraft handling qualities. In effect, this idea touched on the idea of pilot workload. A pilot may be capable of attaining the same performance for a wide range of vehicle characteristics at the expense of capacity to assume other tasks. As a result, significant differences in task performance may not be noticed when very real differences in mission suitability do in fact exist. Hence, pilot evaluation was determined to be important and consisted of pilot ratings and pilot comments. The pilot rating scale was developed as a systematic means of denoting flying qualities of the pilot-vehicle combination in the accomplishment of its intended purpose (mission).

4.2.2. Procedure

The decision-tree format of the Cooper-Harper HQR Scale guides the pilot through a series of dichotomous decisions to arrive at a quantitative value, or rating, for the aircraft handling qualities in the task being performed. The decision process and ratings are illustrated in figure 4-1.



Sequential Pilot-Rating Decisions

Figure 4-1

Figure 4-1 demonstrates the path followed during a handling qualities evaluation using the Copper-Harper HQR. It should be noted that the Cooper-Harper scale is utilized for closed loop tasks, that is where the operator inputs a command, assessing aircraft response, and makes a corrective input based on aircraft motion. This continual input, feedback, input cycle is referred to as closed-loop. One example would be the extremely high gain task of airborne refueling, or tanking, as it is more commonly referred to. In this task, the pilot of the receiver aircraft must fly the airplane into a position below and behind the tanker aircraft so as to place a 6 inch diameter probe into a basket being trailed from the tanker. If during the attempt to fly the probe into the basket the pilot encounters uncontrollable pitch oscillations due to poor control system mechanical characteristics or extremely poor short period response of the aircraft, the Cooper-Harper rating for the in-flight refueling task would be a 10. However, if the same task resulted in aircraft response that was acceptable and satisfactory for the task the Cooper-Harper rating would be 3 or less. It can be seen that within each category in the Cooper-Harper Scale several other definitions should be discussed.

The authors provide that controllable means the pilot must be able to "retain control in all mission-oriented tasks or other required operations without sacrificing effort and attention to . . . overall duties." However, for a level of quality determination, the Cooper-Harper uses the following category definitions:

> Satisfactory is good enough without improvement and therefore of the best category,
>
> Unsatisfactory but tolerable is adequate for the purpose but improvement is desirable,
>
> Unacceptable is not suitable for the purpose but still controllable, and
>
> Uncontrollable is unacceptable for the purpose and of the poorest category.

The evaluator can arrive at one of these four broad categories by considering the three questions represented on the left hand side of the decision-tree:

> Is the vehicle controllable?
>
> Is adequate performance attainable with a tolerable workload?
>
> Is the vehicle satisfactory without improvement?

The Cooper-Harper scale subdivides quality within three of these four categories with appropriate descriptions for each numerical value to further discriminate between HQR ratings. Several key terms are utilized in the final phase of the scale which should be reviewed. Performance is intended to mean the level of achievement attained by the total pilot-vehicle system. Workload refers to the amount of mental and physical effort and attention required by the pilot to attain a given performance level. Compensation is a measure of additional pilot effort and attention required to maintain a given level of performance in the face of deficient aircraft characteristics. The authors present, therefore, that total workload is the workload due to compensation for aircraft deficiencies plus the workload due to the task.

An important consideration is what level of performance is good enough? For that reason, the evaluator must have a firm idea of desired vs. adequate performance levels for each task to be evaluated. Important in the consideration of desired performance is what is the result of not meeting that level. As an example, assume the task under evaluation is pitch control during landing. How precise the operator is expected to capture and hold a nose attitude is dependent on aircraft handling and performance characteristics if those limits are exceeded. An aircraft that handles the same at 4 degrees nose up as it does at 8 degrees nose up during landing may not require precise nose attitude control; however an aircraft that loses all usable lift and develops lateral PIO at 2 degrees above landing attitude would require a tighter performance window to be acceptable. For these reasons, a great deal of attention must be given during the test planning stages to determine what performance window is desired and what is adequate for mission suitability.

4.2.3. Analysis

The Cooper-Harper HQR Scale is the current standard for evaluation of aircraft handling qualities and is currently taught at all three military test pilot schools in the United States and the United Kingdom. Its universal application throughout the aviation test community makes it an excellent starting point for subjective evaluation technique discussion. While the terminology used within the scale refers to "pilot" compensation, the scale could be applied to almost any closed-loop aircraft control task. One of the limitations in applying the Cooper-Harper scale is that the compensation and performance terms are geared towards specific, psychomotor tasks associated with aircraft control. The Cooper-Harper Scale has limited application in the workload measurement arena in advanced system design evaluation. The *American National Standard Guide to Human Performance Measurements* recommends that the Cooper-Harper HQR Scale should be used for workload assessment only if aircraft handling qualities are the major determinant of workload. Current experience dictates that aircraft handling qualities are not normally the driving factor in advanced cockpit system design deficiencies. For instance, the operator's inability to identify, designate, and deliver ordnance on a surface target may have nothing to do with aircraft handling qualities but rather be a function of poor sensor integration and cockpit controls. As a result, it is important that future system evaluators including pilots, WSOs, and flight test engineers be introduced into subjective techniques more suited to workload measurement.

4.3. MODIFIED COOPER-HARPER

4.3.1. Background

The Cooper-Harper HQR Scale combines handling qualities and workload in determining a relative value. Though well known to all test pilots and universally accepted as an excellent standard of measurement for flying qualities evaluations, the Cooper-Harper scale is not well suited for workload measurement purposes largely due to its reliance on psychomotor skills for pilot evaluation, relying on descriptors such as control (aircraft response), compensation (pilot inputs), and performance (quantifiable results). However, the Cooper-Harper Scale has been the model for several workload rating scales which have been subsequently developed. One such scale is the Modified Cooper-Harper Scale shown in figure A-3. The Modified Cooper-Harper

Scale is oriented towards aircrew cognitive functions such as perception, evaluation, and communication and therefore more applicable to modern system evaluation. Table 4-1 lists a comparison of key features of the Cooper-Harper and the Modified scales. One result of the modified rating procedure was a decrease in rating (i.e.- "better" score) associated with minimal effort and adequate performance. For instance, in the Cooper-Harper, the best rating that can be achieved with a result of adequate performance is an HQR 5. However, with the Modified Scale, adequate performance could be reflected by a rating as high as 3. Consequently, it is important to note that Modified Cooper-Harper scores should not be compared "one for one" with Cooper Harper HQR values.

Table 4-1

Comparison between Cooper-Harper HQR and Modified Cooper-Harper Rating Scales

Parameter	Cooper-Harper HQR	Modified Cooper-Harper
Initial Decision	Aircraft Controllability	Frequency of Task Accomplishment
Adequacy for Selected Task	Performance/Workload required	Same
Required Fix	Improve Aircraft Deficiency	Reduce Mental Workload/ Redesign System
Description of Aircraft Characteristic	Highly Desirable to Major Deficiencies	None
Operator Demand Level as Function of	Pilot Compensation	Operator Mental Effort
Endpoints	10 (Highly Desirable - Uncontrollable)	10 (Very Easy - Impossible)

4.3.2. Procedure

Similar to the Cooper-Harper, the Modified Cooper-Harper is designed for use in-flight immediately following the task or maneuver being evaluated. Furthermore, adequate training of the subject evaluator is a must in order to ensure this. operator must be cautioned against the same poor habit patterns that are routinely witnessed when using the Cooper-Harper HQR: premature assignment of a rating score without paying attention to the reference card for key operator demand level discriminators and inconsistent definition of adequate and desired system performance. As with any handling qualities evaluation, a credible system evaluation demands that adequate attention is paid to establishing realistic tolerances on performance criteria prior to commencement of flight tests. However, tolerance levels need not be set in stone once established during test plan development. In fact, some alteration of tolerance levels may be required once flight testing has commenced and system performance with the pilot-in-the-loop has become more clearly defined. For instance, is it really necessary that the pilot be able to designate a the turret of tank-sized target within two feet if the system auto-contrast tracker performance will only provide accuracy within five feet?

4.3.3. Analysis

Various studies since the publication of the Modified Cooper-Harper have indicated that ratings increase as secondary task loading is increased. Secondary tasks include functions such as communications, navigation, and danger conditions. This indicates the Modified Cooper-Harper is indeed a reliable indicator of overall mental workload.

4.4. HONEYWELL COOPER-HARPER RATING SCALE

4.4.1. Background

The Honeywell Cooper-Harper Rating Scale was developed by Wolf (1978) to assess overall task workload. The scale is shown in figure A-4.

4.4.2. Procedure

Unlike the Modified Cooper-Harper Rating Scale, the Honeywell Scale is similar to the Cooper-Harper HQR Scale in terminology. The initial pilot decision is identical: based on aircraft controllability and workload levels required for adequate performance. However, table 4-2 depicts several key differences in the Honeywell Scale that make it more suited to system integration issues.

4.4.3. Analysis

The Honeywell Scale is more suited to an evaluation of aircraft controls or cockpit design that will clearly impact the pilot's ability to control aircraft maneuvers. North, Stackhouse,

Table 4-2

Comparison between Cooper-Harper HQR and Honeywell Cooper-Harper Rating Scales

Parameter	Cooper-Harper HQR	Honeywell Cooper-Harper
Initial Decision	Aircraft Controllability	Same
Adequacy for Selected Task	Performance/Workload required	Same
Required Fix	Improve Aircraft Deficiency	Reduce Workload
Description of Aircraft Characteristic	Highly Desirable to Major Deficiencies	None
Operator Demand Level as Function of	Pilot Compensation	Pilot Effort
Endpoints	10 (Highly Desirable - Uncontrollable)	9 (Pilot effort not a factor - Uncontrollable)

and Graffunder (1979) found the Honeywell scale well suited to assess workload associated with various displays for a Vertical Takeoff and Landing aircraft.

4.5. BEDFORD SCALE

4.5.1. Background

Roscoe and Ellis (1990) reported that in the mid 1970's, advancements in aircraft performance were highlighting the importance of aircrew workload on flight safety. Two examples of this were the advent of Vertical Takeoff and Landing (VTOL) capabilities and low visibility landings. In an effort to support the unstructured pilot opinion, engineers decided on supporting subjective opinions regarding workload by monitoring pilot heart rate. While this physiological technique was reliable, there still existed a need for standardizing pilot comments regarding workload assessment as it impacted aircraft system performance. With no scale currently in existence, Roscoe and Ellis developed the Bedford Workload Scale using current test programs and qualified test aircrew. While the Cooper-Harper decision tree structure would serve as a foundation for the Bedford Scale, the authors determined that a more concrete definition of workload was required. A study of more than 350 military and airline pilots determined that more than 80% of those interviewed thought of workload in terms of effort. This interpretation of pilot workload was relevant in that the "perceived: demands of a flight task would be a function of many factors including training, pilot ability, experience and flight task." Furthermore, pilots indicated it was convenient to use "spare capacity" when trying to measure their workload level. As Roscoe (1990) reported, "Pilots seem to find it a relatively simple matter to judge how much more they could do even if there is no requirement to do so."

As a result, Ellis and Roscoe proposed a new definition: "Pilot workload is the integrated mental and physical effort required to satisfy the perceived demands of a specified flight task."

The Bedford Workload Scale depicted in figure A-5 is a binary decision tree structure based on the ordinal structure of the original Cooper-Harper HQR Scale. A comparison of the Bedford Scale with the Cooper-Harper HQR Scale is presented in table 4-3.

4.5.2. Procedure

Using the Bedford Scale, the evaluator begins the process at the lower left corner and continues through the decision tree in a manner similar to the Cooper-Harper HQR. In fact, the Bedford Scale requires very little additional training once aircrew have had experience with the Cooper-Harper Scale. Half ratings are allowed, which increases the sensitivity of the process by allowing for borderline decisions in some areas. As Roscoe (1990) pointed out, "Originally half ratings between decision groups were not sought but as many pilots seemed to find it difficult to

Table 4-3

Comparison between Cooper-Harper HQR and Bedford Rating Scales

Parameter	Cooper-Harper HQR	Bedford Workload
Initial Decision	Aircraft Controllability	Task Completed?
Adequacy for Selected Task	Performance/Workload required	Same
Required Fix	Improve Aircraft Deficiency	None
Description of Aircraft Characteristic	Highly Desirable to Major Deficiencies	None
Operator Demand Level as Function of	Pilot Compensation	Spare Capacity
Endpoints	10 (Highly Desirable - Uncontrollable)	WL1 (Workload Insignificant) - WL10 (Task Abandoned)

decide between 'yes' and 'no' for 'Was workload satisfactory without reduction?' a rating of 31/2 became acceptable."

Pilot training is necessary to ensure all personnel are familiar with differences in rating levels and standardized terminology usage. the subject flight task must be well defined with a set time limit over which the assessment should be conducted. Also, as with other workload assessments, the test team must ensure that it is only task loading associated with the primary task that is being evaluated. Roscoe points out that any other additional (secondary) tasks to be conducted must be considered as part of pilot "spare capacity" when using the scale.

4.5.3. Analysis

As with other subjective rating scales, the Bedford Scale was designed for use in-flight. The first use of the rating scale was during a Harrier Ski-Jump Take Off trial which utilized the Cooper-Harper HQR Scale for flying qualities and the Bedford rating Scale for workload level. Heart rates were recorded during the evaluation by 11 different pilots in an attempt to substantiate the ratings assigned using the Bedford Scale. Roscoe (1984) reported that the "workload ratings and heart rates showed good agreement and both confirmed that workload levels were not increased for greater ramp angles or night takeoffs."

Lidderdale (1987) used the Bedford Scale during a low-level evaluation in a tactical fighter aircraft and reported that the aircrew "understood the scale readily and whilst it was sufficiently comprehensive to cover all circumstances it was easy to remember..." Lidderdale also conducted the post-flight Analytical Hierarchy Process (a mathematical analysis using paired combinations to determine workload levels) to determine relative worth of the Bedford Scale. While determining there was a high correlation between in flight ratings and the post flight assessment, Lidderdale observed that the Bedford Scale was "easier to use and more practical for use in an operational trial." Roscoe (1984).

The primary advantage to the Bedford Scale is its simplicity and applicability to in-flight evaluations. Roscoe and Ellis (1990) reported that they are "unaware of any other workload rating scale that has been used as extensively for assessing workload in the real world."

4.6. ALTERNATIVES

The use of rating scales has been questioned in several areas. There is a lack of sensitivity in certain regions of the rating values and limited diagnostic capability when the only post-flight data is a rating generated by aircrew opinion. There is only a relative, or ordinal, nature to the rating. Some studies have been generated to develop a more absolute rating system, though as Lidderdale (1987) pointed out, due to the multi-dimensional nature of workload assessment, all measurements should be considered relative. There are several techniques in use, however, which solve some of the problems through use of a different approach.

4.7. DIAGNOSTIC TECHNIQUES

4.7.1. Subjective Workload Assessment Technique

4.7.1.1. Background

While rating scales do provide good in-flight measurement tools, they are ordinal in nature; that is they provide some idea of the order or relativity of workload associated with a specific flight task. As a result, any statistical analysis capability is severely limited. In order to develop a more absolute measure of workload, the Subjective Workload Assessment Technique (SWAT) was introduced in the early 1980's. Generally, it is accepted that no single dimension can describe workload; as a result SWAT defines workload in terms of three dimensions as presented in figure A-6. The SWAT combines ratings on these three scales to produce a diagnostic value of operator workload.

4.7.1.2. Procedure

SWAT is based on conjoint measurements and scaling and allows ratings on the three dimensions to be combined into one overall workload scale. In order to identify the appropriate mathematical process for combining the three dimensions into one overall scale, a "scale development" phase must be completed. (Schick and Hann, 1987) A complete description of the SWAT mathematical analysis can be found in Reid (1982).

4.7.1.3. Analysis

Schick and Hann (1987) conducted a validation experiment consisting of a series of landing approached flown in a cockpit simulator using a German language version of the published SWAT. The report indicated that SWAT ratings showed "highly significant differences between the individual pilots" and an increasing in SWAT ratings which corresponded to an increase in task difficulty levels. The report concluded that "basic evidence was found for the validity of SWAT as a tool for the assessment of mental workload , and thus for its applicability in . . . actual in-flight tests."

Schick and Hann reported that the SWAT data collection must be carefully planned so as not to interfere with flight tasks. They recommend against planning to conduct data collection at specific time intervals. Rather they recommended that "SWAT data collection should be organized as being event-related, in a way which assures that both subject and experimenter have the same understanding as to which task or sequence of tasks . . . a SWAT rating refers to."

Not surprisingly, this technique has limited application in the school environment; it is however, one of the more common statistical analysis methods in industry today and should be introduced to the student.

4.7.2. NASA TASK LOAD INDEX

4.7.2.1. Background

The NASA Task Load Index (TLX) was developed by the Human Performance Group at the NASA Ames Research Center at NAS Moffet Field, CA and defines workload as the "cost incurred by human operators to achieve a specific level of performance." (NASA TLX Pencil and Paper Package v1.0) The TLX was developed as a follow-on to the NASA Bipolar Rating Scale which was found to have insufficient sensitivity to experimental manipulations and contained too many subscales, making the scale impractical to use in a simulation or operational environment. The subjective experience of workload is defined by a combined effect of subjective responses (emotional, physical, and cognitive) and a series of weighted behavior evaluations. The process is requires more time to comprehend than a rating scale approach but is more suited to a statistically analysis. An experimentally based process of elimination led to the identification of six dimensions for the subjective workload experience:

SUBJECT ONLY	SUBJECT/TASK INTERACTION
Mental Demands	Performance
Physical Demands	Effort
Temporal Demands	Frustration

Unlike most unidirectional rating scales, the TLX can help in diagnosing which of the six dimensions contributed to a specific task workload. The overall workload of a particular task is determined by weighting each individual factor, thus increasing scale sensitivity with respect to each dimension's role in the task loading of a test maneuver.

As an illustration, consider two pilots, A and B. Pilot A and B are told that there is a contact at 230 degrees at 21 nautical miles. Pilot A has trouble manipulating the cursor controller (physical task), pilot B can not read the navigation display (mental task). The same task, designating the airborne threat, is difficult for both pilots but for very different reasons. Whereas a rating scale may not account for this, the TLX identifies the problem through use of the 6 dimensions.

4.7.2.2. Procedure

The evaluation should be based on both performance and workload perception. (Just because the evaluator was successful at locking a target within the desired range does not warrant a high score; possibly it took all work capacity to perform within pre-determined limits.) Rating scale definitions, as shown in figure A-6 must be clearly understood.

Use of the TLX requires two steps. First, subjects rate each task performed on each of the six subscales shown in figure A-7. Second, the subjects must perform 15 pair-wise comparisons of six workload scales which are diagrammed in figure A-8. The number of times each scale is rated as contributing more to the workload of a task is used as the weight for that scale. This step determines the relative importance of each dimension, since each pilot is certain to have differing impressions on what dimensions play into the workload requirement of any particular task. For instance, some evaluators may feel that mental or temporal demands are the important aspect of workload regardless of effort level or performance achieved. Another evaluator may bias the rating scales towards performance: radar lock and missile fire within thirty seconds means that workload must be low. This relative weighting is accomplished through use of the Workload Comparison Card, in figure A-9.

4.7.2.3. Analysis

Hart and Staveland (1987) concluded that the TLX provides a sensitive indicator of overall workload as it differed among tasks of various cognitive and physical demands and that the TLX provided valuable diagnostic information about the sources of workload in any particular task. This can be a lengthy procedure, as each comparison appears on a separate card requiring each task to be followed by a flip through 15 cards. The data reduction is quite complex; computer software support is available from NASA Ames Research Center, CA.

5. CONCLUSION

5.1. STUDENT SYLLABUS

In order to offer a more advanced approach to qualitative assessment of airborne systems design, a student program of instruction was implemented in Class 105 at the USNTPS in February, 1994. As previously discussed, the syllabus was modified to increase student exposure to subjective assessment techniques and pilot workload consideration in system evaluation exercises. As discussed in Section 2, three Advanced Systems Intigration (ASI) flight exercises (ASI I, II, and III) constitute the student test pilot airborne system exposure while at USNTPS. ASI I introduces the student to human factors considerations in the evaluation of an unfamiliar air-to-ground radar system. ASI II requires the student to conduct an evaluation of an unfamiliar air-to-air radar system with a focus on subjective workload assessment to validate aircrew opinion regarding system integration.

At USNTPS the primary airborne systems aircraft is the McDonnel Douglas F/A-18 Hornet. The simulator demonstration is conducted in the Manned Flight Simulator (MFS) motion-based dome simulator at NAS Patuxent River. (In order that all students see an unfamiliar system, students with prior F/A-18 experience are put through the same syllabus in an aircraft like the F-14D or F-16B.)

The Advanced Systems Integration Program of Instruction is outlined in table 5-1. Students receive three hours of classroom instruction prior to the simulator demonstration. The entire program is set up to give each student exposure to the entire spectrum of a systems evaluation form test plan development through flight testing to reporting results to program sponsors. Specific periods of instruction are detailed in the following paragraphs.

Table 5-1

Advanced Systems Integration Program of Instruction

Title	Academic/Flight Hours (per student)	Media
Pilot Workload	1.0	Classroom
Performance Measurement	1.0	Classroom
Exercise Brief	1.0	Classroom
Simulator Demonstration	1.5	Manned Flight Simulator
Flight Evaluation	1.5	F/A-18B
Event Debrief	1.0	Classroom

5.2. PILOT WORKLOAD

Students are introduced to the need for credible evaluation techniques to quantify pilot opinion as outlined in Section Three. The brief includes an overview of Air Force Crew Station Evaluation methodology to assist in planning for the ASI II Exercise and systems integration effects in general. Systems effectiveness is introduced as a function of three factors: performance, workload, and situational awareness. Workload is identified as the emphasis in the ASI II Exercise in order to introduce the students to various existing techniques available to test aircrew. The student is instructed on subjective assessment techniques as presented in Section 4. The first iteration of this program emphasized the Modified Cooper Harper and Bedford Scales due to their ease of use and limited training required for implementation.

5.3. PERFORMANCE MEASUREMENT

The second period of instruction introduces the student to the MOP's presented in the USNTPS Systems Test Manual, reference (1). Classroom instruction highlights flight test techniques and data collection considerations. Students are instructed on relationship of system performance and mission suitability issues.

5.4. EXERCISE BRIEF

The Student Exercise Planning Guide, Appendix C, serves as the main student planning document. Students are grouped in test teams: each team consists of one Project Officer and several Project Pilots. If possible, each Project Officer will be an aviator with fleet air-to-air experience.

A key facet to the ASI II exercise is the requirement for the students to develop the mission suitability requirements for the system to be evaluated. The Airborne Systems Staff at USNTPS has identified several key objectives with this:

a. Individual thought on the part of the student is encouraged,

b. Eliminates the "spoon feeding" nature of the exercise, and

c. Students tend to develop a more credible approach when faced with defending their own test maneuvers to their peers.

5.5. SIMULATOR DEMONSTRATION

The simulator period provides build-up for techniques and maneuvers to be conducted during the Flight Evaluation. For most students, the simulator is the first experience with MOW and MOP techniques discussed during the two academic periods. Test Plans are due four days after the simulator demonstration to give each team the opportunity to assess the evaluation system prior to finalizing the test matrix.

Students are expected to brief MOP and MOW techniques required during flight tests. MOP's should be conducted in accordance with reference (1) and MOW's using the appropriate lecture notes for the Modified Cooper Harper or Bedford Scale. Specific Measures of Situational Awareness (MSA's) are beyond the scope of this exercise but students are encouraged to discuss various techniques to evaluate an aircrew's SA (unusual attitudes, questions from ground controllers, navigational errors, etc.)

5.6. FLIGHT EVALUATION

The Flight Evaluation should mirror the profile flown in the simulator as much as possible. Certain simulator limitations will require that some in-flight maneuvers can not be rehersed prior to flight testing. The operational maneuvers are flown to evaluate workload assessment and rating scales are used real time to avoid the memory loss associated with post-flight techniques.

Post-flight questionnaires are completed and used to support flight test ratings. Data reduction will include averaging pilot workload measurements and reporting results as required to support mission suitabilty determination.

5.7. SUMMARY

Increasing system capabilities and the importance of aircrew opinion dictate that future airborne system evaluations must include a measure of operator subjective assessment. At USNTPS, the Airborne Systems curriculum provides an ideal platform and student base with which to implement the Advanced Series Integration II program of instruction outlined herein. If given the proper orientation on the importance of subjective evaluation, how it fits in to the system effectiveness equation, and current techniques available for use, the student evaluator will be given a much better foundation on which to build in the test and evalaution community.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abraham, R. MAJ, USAF. (1994) Personal conversation.

Casali, J.G. and Wierville, W.W. (1983) "A Comparison of Rating Scale, Secondary Task, Physiological, and Primary-task Workload Estimation Techniques in a Simulated Flight Task Emphasizing Communications Load." Human Factors Vol. 25, 623-642.

Cooper, G.E. and Harper, R.P. (1969) "The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities." NASA TN D-5153.

Ellis, G.A. and Roscoe, A.H. (1985) "The Airline Pilot's View of Flight Deck Workload: A Preliminary Study Using a Questionnaire." Royal Aircraft Establishment Technical Memorandum No. FS(B)465.

Guide to Human Performance Measurements. American National Standards Institute/American Institute of Aeronautics and Astronautics G-035-1992.

Harper, R.P. and Cooper, G.E. (1984) "Handling Qualities and Pilot Evaluation," Presented as Paper 84-2442 at the AiAA/AHS/ASEE Aircrat Design, Systems, and Operations Meeting, San Diego, CA.

Hart, S.G. and Staveland, L.E. (1987) "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research." In P.A. Hancock and N. Meshkati (Eds) Human Mental Workload. Amsterdam: Elsevier.

Johannsen, G., et al. (1979) "Final Report of the Experimental Psychology Group," In N. Moray (Ed) Mental Workload: Its Theory and Measurement. New York, Plenum Press.

Landmann, M.A., et al. (1992) United States Naval Test Pilot School Systems Test Manual, Airborne Systems Theory and Flight Test Techniques.

Lehman, et al. (1989) "Handbook for Conducting Pilot-In-The-Loop Simulation for Crew-Station Design." USAF Systems Command HSD-TR89.

Lidderdale, I.G. (1987) "Measurement of Aircrew Workload During Low-level Flight." In: Practical Assessment of Pilot Workload. Roscoe, A.H. (ed). AGARDograph No. 282., AGARD Paris.

Milam, D.W. (1982) "A PIlot's Perspective of Workload in Single Seat Fighters." In: Proceedings to the Workshop on Flight Testing to Edentify Pilot Workload and Pilot Dynamics, AFFTC-TR-82-5.

NASA Task Load Index (TLX) Paper and Pencil Package (v1.0). Human Performance Research Group, NASA Ames Research Center, Moffet Field, CA.

North, R.A., Stackhouse, S.P., and Graffuder, K. (1979) "Performance, Physiological and Oculometer Evaluations of VTOL Landing Displays." NASA Langley Research Center Contractor Report 3171, Hampton, VA.

O'Donnell, R.D. and Eggemeir, F.T. (1986) "Workload Assessment Methodology." In: K.R. Boff, L. Kaufman, and J.P. Thomas (Eds.) Handbook of Perception and Human Performance. New York: Wiley and Sons.

O'Lone, R.G. (1992) "Boeing Works to Keep Pilot in Decision Loop On Board Advanced 777." Aviation Week & Space Technology, 23 Mar. 1992: 61.

Reid, G., Eggemeier, F., and Shingeldecker, C. (1982) "Subjective Workload Assessment Technique." Presented at the 1982 Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics.

Roscoe, A.H. (1984) "Assessing Pilot Workload in Flight" In: Flight Test Technique. AGARD Conference Proceedings No. 373. AGARD, Paris.

Roscoe, A.H. (Ed) (1987) "Subjective Workload Assessment Techniques." In: Practical Assessment of Pilot Workload. Roscoe, A.H. (ed). AGARDograph No. 282., AGARD Paris.

Roscoe, A.H. and Ellis, G.A. (1990) "A Subjective Rating Scale for Assessing Pilot Workload in Flight: A Decade of Practical Use." Royal Aerospace Establishment TR 90019.

Schick, F.V. and Hann, R.L. (1987) "The Use of Subjective Workload Assessment Technique in a Complex Flight Task." In: Practical Assessment of Pilot Workload. Roscoe, A.H. (ed). AGARDograph No. 282., AGARD Paris.

Sparaco, P. (1987) "Human Factors Cited in French A320 Crash." Aviation Week & Space Technology. 3 Jan. 1987: 30.

Spear, J. LT, RAF. (1994) Personal conversation.

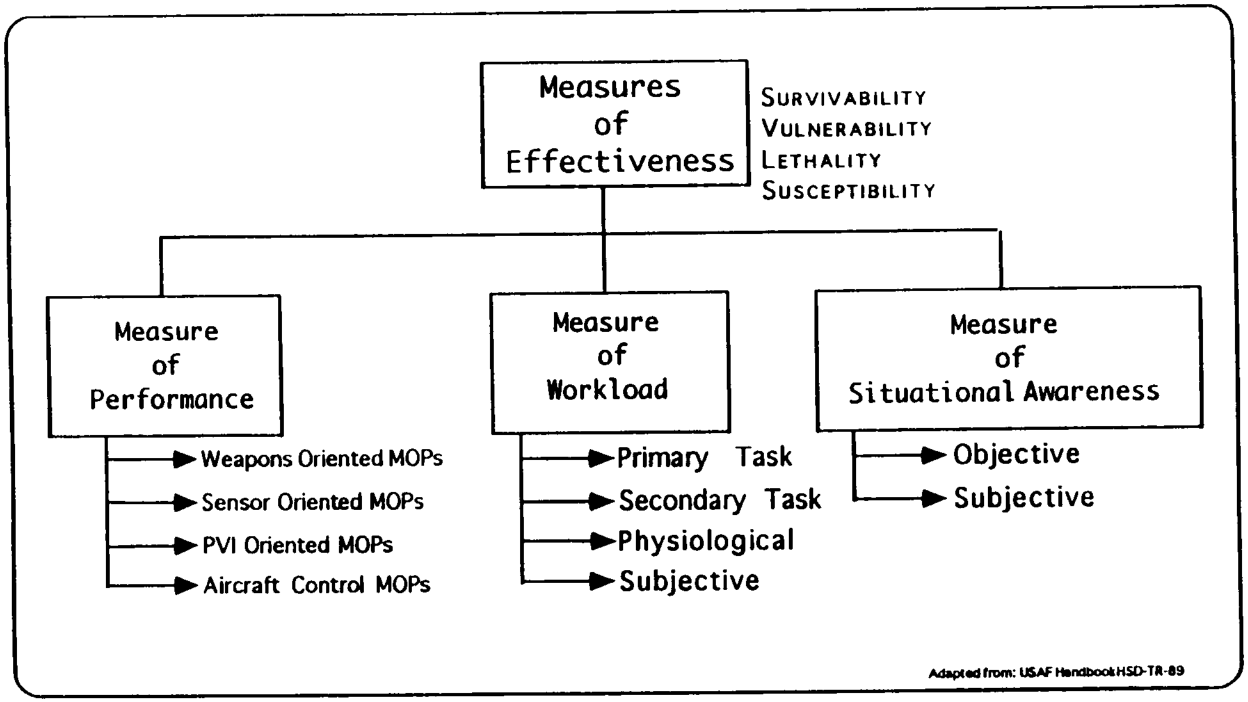
Vidulich, M.A. (1991) "The Bedford Scale: Does it Measure Spare Capacity?" In: Procedings of the Sixth International Symposium on Aviation Psychology: April 29-May 2, 1991.

Wierville, W.W. and Casali, J.G. (1983) "A Validated Rating Scale for Global Mental Workload Measurement Applications." In: Proceedings of the 27th Annual Meeting of the Human Factors Society. Human Factors Society, Santa Monica, CA.

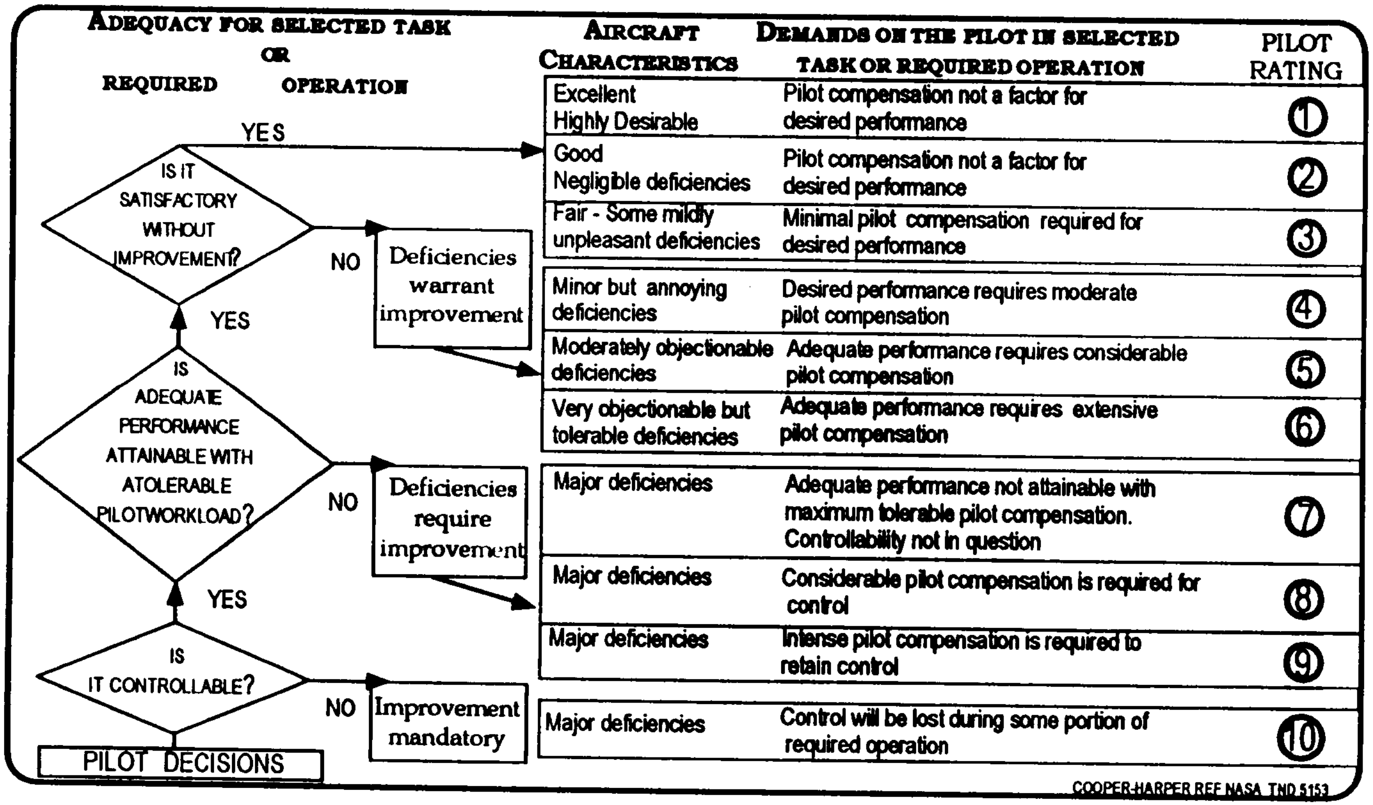
Wolf, J. D. (1978) "Crew Workload Assessment: Development of a measure of Operator Workload." Air Force Flight Dynamics Labratory TR-78-165. Wright-Patterson AFB, OH.

Woodfield, A.A., and Renaudie, J.F. (1987) "The Practical Assessment of Pilot Workload." In: Practical Assessment of Pilot Workload. Roscoe, A.H. (ed). AGARDograph No. 282., AGARD Paris.

APPENDIX A: FIGURES

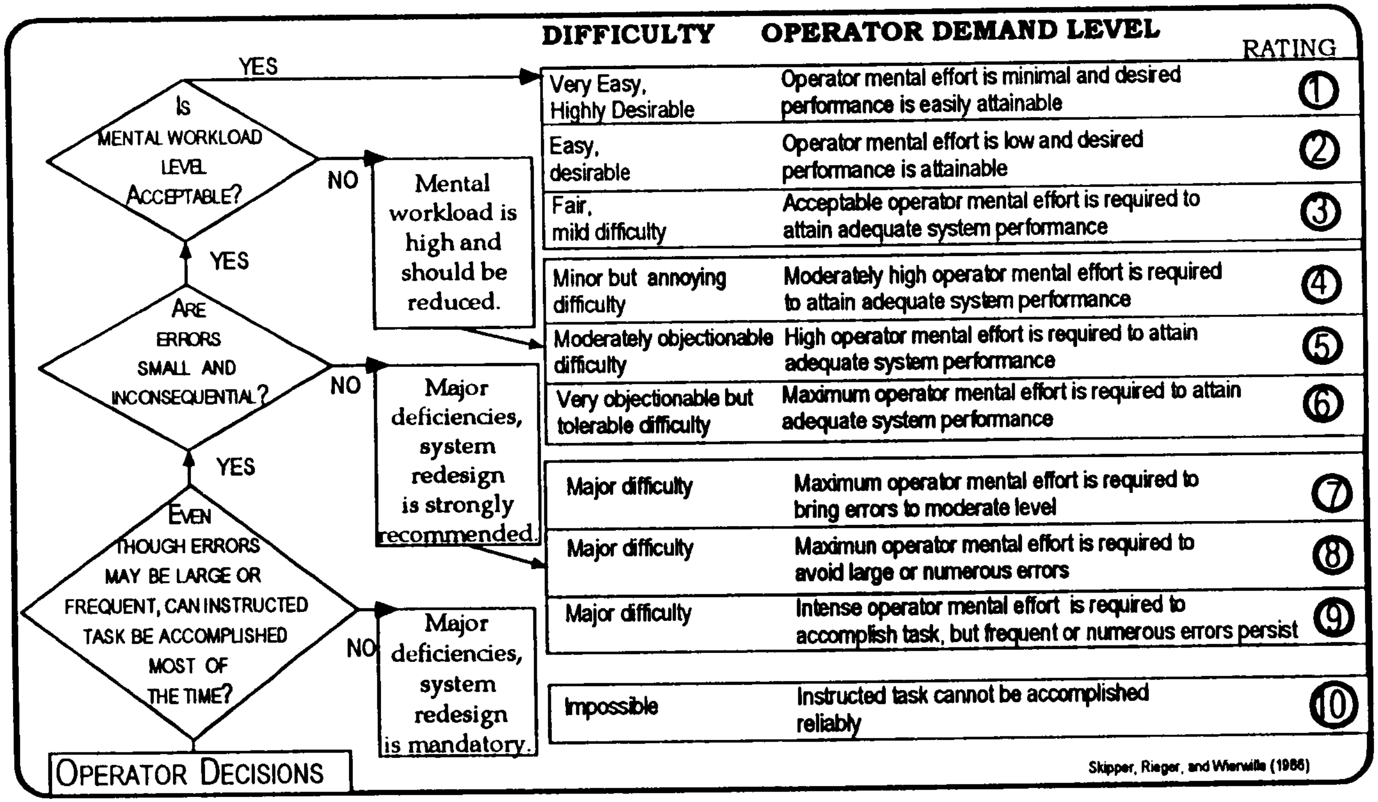


Hierarchy of Evaluation Measures
Figure A-1



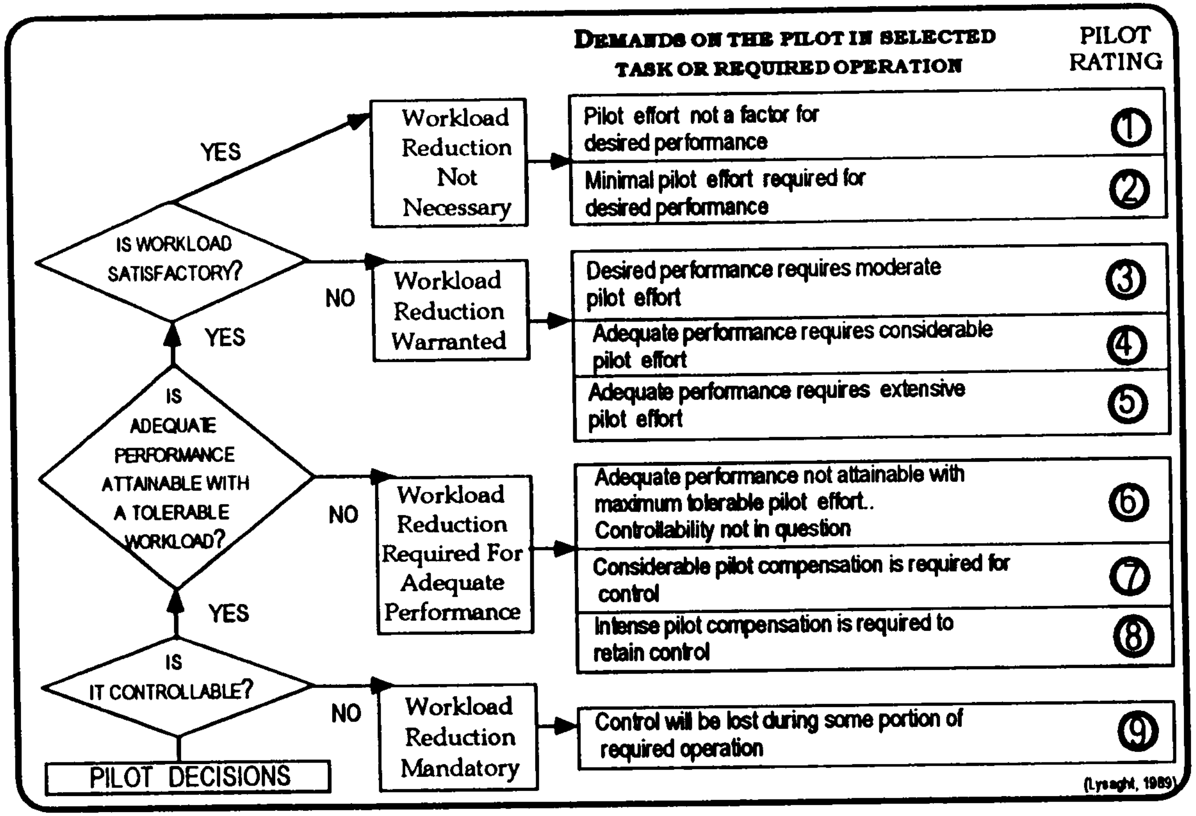
Cooper Harper Handling Qualities Rating Scale

Figure A-2



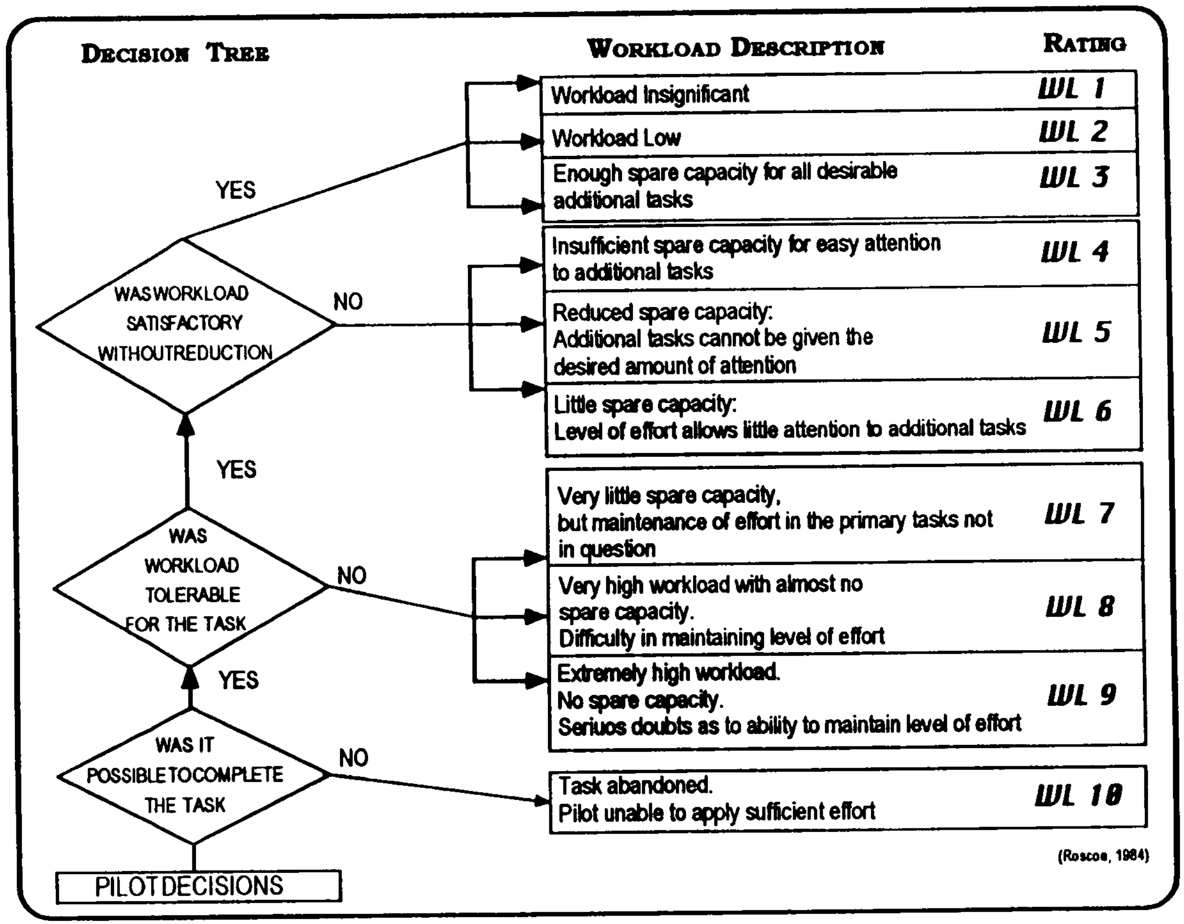
Modified Cooper-Harper Ratings Scale

Figure A-3



Honeywell Cooper Harper Ratings Scale

Figure A-4



Bedford Workload Scale
Figure A-5

The Three SWAT Dimensions	
LEVELS	DIMENSION
	I. TIME LOAD
1	Often have spare time. Interruptions or overlap occur infrequently or not at all.
2	Occasionally have spare time. Interruptions or overlap among activities occur frequently.
3	Almost never have spare time. Interruptions or overlap among activities are very frequent, or occur all the time.
	II. MENTAL EFFORT
1	Little conscious mental effort or concentration required. Activity is almost automatic, with little or no attention.
2	Moderate conscious mental effort or concentration required. Complexity of activity is moderately high due to uncertainty, unpredictability, or unfamiliarity.
3	Extensive mental effort and concentration are necessary. Very complex activity requiring total attention.
	III. PSYCHOLOGICAL STRESS
1	Little confusion, frustration or anxiety exists and can be easily accommodated.
2	Moderate stress due to confusion, frustration or anxiety. Noticeably adds to workload. Significant compensation is required to maintain adequate performance.
3	High to very intense stress due to confusion, frustration or anxiety High to extreme determination and self-control required.

Swat Table

Figure A-6

RATING COMPARISION CHECKLIST

(CIRCLE AS APPROPRIATE)

Effort or Performance	Temporal Demand or Frustration
Temporal Demand or Effort	Physical Demand or Frustration
Performance or Frustration	Physical Demand or Temporal Demand
Physical Demand or Performance	Temporal Demand or Mental Demand
Frustration or Effort	Performance or Mental Demand
Performance or Temporal Demand	Mental Demand or Effort
Mental Demand or Physical Demand	Effort or Physical Demand
Frustration or Mental Demand	

NASA TLX Scale
Figure A-7

RATING SCALE DEFINITIONS

TITLE	ENDPOINT	DESCRIPTIONS
Mental Demand	*Low/High*	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex exacting or forgiving?
Physical Demand	*Low/High*	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand	*Low/High*	How much time pressure did you feel sue to the rate or pace at which the tasks or elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	*Perfect/ Failure*	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Effort	*Low/High*	How hard did you have to work (mentally and physically) to Accomplish your level of performance?
Frustration Level	*Low/High*	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

Figure A-8

APPENDIX B: SAMPLE WORKLOAD TASK

SAMPLE WORKLOAD TASK

This five minute recce/attack task for fast jet aircraft is adapted from Appendix 2 of AGARDograph No. 282 and demonstrates the level of detail commonly associated with preflight planning of workload assessment flight maneuvers.

1. CHOICE OF TARGET FOR RECCE/ATTACK

A realistic target would be a simple, soft-skinned vehicle parked on a two-lane track at a known position in flat/undulating open country.

2. ROUTE PARAMETERS FOR ATTACK TASK

Route comprises a single 3 minute navigation leg to Initial Point (IP) followed by 1 minute run to the target for a visual delivery utilizing a standard pop-up to a visual dive delivery. Aircraft pop-up maneuver is commenced at the Pop-Up Point, and a shallow dive delivery is established through the Weapon Release Point (WRP): specific delivery profile will depend on aircraft type. Following attack, defensive recovery maneuver and fix before egress 1 minute after target. (Recce task would be similar sequence but camera selection and offset fly-by would replace weapon delivery and tip-in attack.)

3. ACCURACY

En route and IP run: $\pm$ 20 knots/ $\pm$ 100 feet

Pull-up point: $\pm$ 1000 feet laterally, $\pm$ 3 seconds

Dive angle: $\pm$1 degree

Release point: $\pm$20 knots

4. DETAILS OF ATTACK TASK

a. During approach to IP, complete the following:

Trim out sideslip at planned release speed

Complete weapons checks

1x track check in air-to-ground map

1x revision estimated time of arrival for IP $\pm$ 5 seconds

8x simulated check of wingman aircraft 6 o'clock position

Determine wind speed/direction (estimate or INS) and set mil depression for back-up attack

Carry out height fix at IP to update pressure altitiude information for weapons release

Set next heading

Acquire IP visually

b. IP to Pull-Up Point (PUP)

Complete height fix (if INS equipped)

Verify position/update INS (if required)

Accelerate to attack speed

Check track to PUP and PUP time

Adjust sight picture for weapons release

c. PUP to WRP

Initiate Pull-Up

Acquire target

Top at height required to achieve desired dive angle

Check speed

Sight or bomb fall line on target

Final arming switchology

Start camera

Track target for 3 seconds prior to WRP

Release weapon

d. Escape

Recover from dive

Make defensive maneuver

Safe weapon switches

Track to egress point

Locate wingman

Stop camera

Regain HUD navigation mode

APPENDIX C: STUDENT EXERCISE PLANNING GUIDE

ADVANCED SYSTEMS INTEGRATION
STUDENT EXERCISE PLANNING GUIDE

1. PURPOSE

1.1. General

1.1.1. Introduce student to subjective workload assessment techniques.

1.1.2. Practice ability to clearly identify mission suitability criteria prior to flight test planning and execution.

1.1.3. Demonstrate a working knowledge of system performance test techniques.

1.1.4. Increase flight test experience to include an advanced air-to-air weapon control system.

1.1.5. Gain exposure to state-of-the-art air-to-air weapons systems and discuss potential improvements to problems identified during flight testing.

1.1.6. Demonstrate ability to propose hardware, software, or functional/procedural change to resolve identified weapon system deficiencies.

1.1.7. Deliver an informal team debrief of flight test results.

1.2. Specific

1.2.1. Evaluate the assigned aircraft radar system and associated equipment for mission as assigned herein.

1.2.2. Demonstrate ability to evaluate system effectiveness as determined by appropriate Measure Of Workload and Measure Of Performance flight test techniques. MOW techniques would include one in-flight (Bedford or Honeywell Cooper-Harper) and one post-flight technique (questionnaire). MOP techniques will be as listed in reference (1).

1.2.3. Demonstrate flight test techniques described in Section III-C of reference (1), to include at least eight of the following MOP's:

Antenna Stabilization	Azimuth Resolution
Antenna Scan Rates	Velocity Resolution
Antenna Scan Angle Limits	Minimum Range
Elevation Angle Limits	Probability of Detection
Antenna Tracking Rate Limits	Maximum Detection Range
Blind Velocities	Blind Ranges
Range Resolution	Bearing/ Range Accuracy

1.2.4. Develop techniques for incorporating simulator assets into test plan for hazard analysis and cost reduction.

1.2.5. Experience advanced cockpit design and system capabilities like HOTAS, digital displays, and electro-optic sensor integration.

1.2.6. Practice Quick Response Report procedures.

2. AIRCRAFT/INSTRUMENTATION:

2.1. Simulator portion will be conducted during one three hour simulator period per test team. No special instrumentation is available.

2.2. Flight event will be conducted during one 1.5 hour flight in assigned aircraft. TSTR-01 has no data recording instrumentation installed; TSTR-02 has capability for 60 minutes of over-the-shoulder (OTS) video for HUD and DDI recording. F-14D players will need to confirm instrumentation available with Strike F-14 Project Office POC, LT Dave "Cheese" Asjes, x-3982.

3. TEST PLANNING:

3.1. Test team assignments are listed in table C-1:

Table C-1

Test Team Assignments

Project Officer	A/C	Project Pilots	Mission	Grader
LCDR Capstaff	F-14D	Frick	Fleet Air Defense	CDR Clark
MAJ Heins	F-14D	Drechsler/Baker	Long Range Interceptor	LCDR Rabens
Plumb	F/A-18B	Dishman/Kelly	Self Escort	Wilmore
Kelly	F/A-18B	Short/Kumke	Point Defense Fighter	Bowers
Gassie	F/A-18B	Waurio/Trost	Fleet Air Defense	Bowers

3.2. Each Test Team will prepare one test plan as detailed below and submit to the Exercise Monitor within 4 days following completion of the simulator demonstration period.

3.#.. Begin test planning process by deciding, as a group, what criteria are necessary to ensure mission suitability. For the purposes of this exercise, mission suitability should be considered as a combination of system performance and system integration - to be measured using MOP's and MOW's as discussed in the exercise brief. Each team must list the following items in the Test Plan:

3.3.1. Eight MOP's (minimum) to be evaluated from the list in paragraph 1.2.2 above. The list of MOP's is at the discretion of the Test Team and should be based on mission requirements.

3.3.2. Three operational maneuvers (minimum) which will be flown in order to assess system integration/crew station design as it applies to pilot workload. Each Test Team will be required to demonstrate one in-flight and one post-flight MOW technique as discussed in the exercise brief.

3.4. Criteria for mission suitability, as determined by the Test Team, should be listed under "TEST STANDARDS" in the "SCOPE OF TESTS" section as described in Enclosure (1) of reference (2). What this means is that the Team must determine what parameters are required to fulfill the assigned mission and then list these parameters in the Test Plan. This will permit the Team to compare flight data against some hard requirement to determine if the aircraft is mission suitable (the artificiality here is that the Team is, in effect, writing the specification). This portion of the Test Plan must be considered very carefully: each MOP and MOW should have a required value for the test/maneuver to be performed. (i.e.- P_d .9 against 5 square meter target not less than __ miles, or successful acquisition of fighter size target and deliver accurate BRAT report within 8 seconds of initial detection (P_d .5). Remember, performance criteria must be quantifiable; MOP's are by nature, MOW's will take some thought. Each team will need to discuss and then list its own criteria for maneuver success (much like the tolerance level on an HQR.)

3.5. Operational Maneuvers to be performed must be specified and detailed in the Detailed Method of Test (DMOT) section. This is similar to what would be diagrammed on a briefing board or on a kneeboard card. Standard air-to-air diagrams apply: who the players are, what the set-up is, and what the expected flight profile will be. Include pertinent "knock-it-off" criteria or other safety of flight information.

3.6. Review Reference (1) and be prepared to demonstrate knowledge of test techniques for MOP's. Again, emphasis should be placed on those parameters with greatest impact on mission requirements as identified in "TEST STANDARDS" section of test plan.

3.7. Each project pilot will be required to perform the three operational maneuvers as depicted in the DMOT. The Test Team will be able to evaluate results from each pilot conducting the same maneuver to evaluate the variance in workload assessment.

4. PREPARATION:

4.1. Simulator Demonstration

4.1.1. The test plan should be well-developed prior to the simulator demo. The test plan is not due until four days after the simulator so any maneuvers or problems identified in the demo can be used to modify the test plan as necessary.

4.1.2. Data in the simulator is valid, though must be supported by in-flight results.

4.1.3. Plan on simulator target aircraft position to be the same on all intercepts (45 miles on the nose, 1000' above ownership on reciprocal heading.) The sim is not a fleet trainer and changing those parameters may be difficult. Don't expect a maneuvering target in the sim.

4.2. Flight Evaluation

4.2.1. Turn in completed test plan to Exercise Monitor within 4 days of simulator demonstration.

4.2.2. Debrief test plan with Exercise Instructor prior to first data flight. Teams will be scheduled together in blocks so anticipate all team members flying evaluation in quick succession.

5. BRIEF:

5.1.. Simulator Demonstration: Plan to be at Manned Flight Simulator (MFS) 30 minutes prior to simulator time. IP will brief admin items and Project Officer will brief MSA test techniques.

5.2. Flight Evaluation: Student will brief flight conduct; IP will brief emergency procedures and safety of flight issues.

6. EXERCISE EVENTS:

6.1. Simulator Demonstration

6.1.1. Each team will be assigned for one 3.0 hour simulator period. Names/simulator period will appear on flight schedule. Please be on time to avoid costly delay at MFS.

6.1.2. The exercise will require three F/A-18 sim periods at MFS. (Expect the week of 28 MAR.) F/A-18 sim periods will have one fleet experienced Hornet student assigned to assist with weapon employment and mission relation issues. IP will be present to discuss flight test techniques and answer any other questions. Students will have the remainder of the 3.0 hour period to utilize as desired.

6.1.3. F-14D sim period currently scheduled 23 MAR at NAS Miramar (0800-1100) Anticipate early departure from EAFB morning of 23 MAR to NAS Miramar, fly sim, and return for airlift to St. Louis.

6.1.4. Simulator profile is at discretion of test team. IP will be available as long as necessary to demo test techniques and answer questions regarding flight procedures.

6.2. Flight Evalaution

6.2.1. Flight events will consist of test maneuvers listed in approved test plan. Flight time is limited to 1.5 hours per student.

6.2.2. Anticipate F-14D flights with SATD sometime mid-April.

7. REPORT

7.1. A Quick Response Report (QRR) will be submitted by the Project Officer to the Exercise Monitor within 7 days of the last test flight.

7.2. IP will snivel Project Team for an Informal Debrief following review of QRR. Debrief will simulate departmental discussion with Department Head and Chief Engineer.

Presentation should cover the basics and be prepared to address other issues as required. Attached Grade Sheet will assist in preparation.

7.3. At a minimum, each Team should address mission suitability requirements as listed in Test Plan. MOP data should be presented; only discussion required is those MOP's that impact mission suitability.

7.4. Plan to discuss "lessons learned" regarding in-flight and post-flight MOW techniques with staff members. This exercise will continue to grow based on student inputs.

8. REFERENCES:

(1) USNTPS Systems Test Manual (Preliminary), 18 JUN 92, Section III-C.

(2) "Preparation, Review, and Implementation of Project Test Plans," FTEGINST 3960.1, dtd 14 JUL 1992.

(3) USNTPS Airborne Systems Course Textbook, 01 FEB 81, pp. 2.34-2.47

(4) Introduction To Airborne Radar, Stimson, G.W., Hughes Aircraft Company 1983.

VITA

Captain Craig Shepperd Bowers was born in Scottsdale , Arizona on April 2, 1961. He graduated from Rocky River High School in Rocky River, Ohio in June, 1979 and entered Georgia Institute of Technology the following September. He graduated in June, 1983 with a Bachelor's Degree in Mechanical Engineering and was commissioned as a Second Lieutenant in the United States Marine Corps. After attending The Basic School in Quantico, Virginia, he transferred to Pensacola, Florida for flight training. He received his wings on February 14, 1986 and was transferred to Whidbey Island, Washington for training in the A-6E Intruder attack aircraft. In October, 1986, he was transferred to Marine Corps Air Station El Toro, California, where he served 4 years flying A-6E's with the United States Navy aboard the aircraft carrier USS Ranger. In June, 1990 he was transferred to Naval Air Station Patuxent River, Maryland to attend Class 99 at the United States Naval Test Pilot School. Upon graduation in June 1991, Captain Bowers was transferred to Strike Aircraft Test Directorate to serve as a Test Pilot on the T-45 trainer aircraft program. He was transferred to the staff at Test Pilot School in May 1992 where he currently serves as Aviation Safety Officer and Instructor Test Pilot in the Airborne Systems.

Captain Bowers has logged more than 300 carrier arrested landings and over 2300 total flight hours in over 30 aircraft types including fixed wing, rotary wing, and glider aircraft. His airborne systems experience includes the AV-8, F/A-18, F-16, F-14, F-15E, and numerous other tactical aircraft.