NASA Technical Memorandum 103947

USAATCOM Technical Report 92-A-006

Test Techniques for Evaluating Flight Displays

Loran A. Haworth, Aeroflightdynamics Directorate, U.S. Army Aviation Systems Command,
Ames Research Center, Moffett Field, California
Richard L. Newman, Crew Systems, San Marcos, Texas

February 1993

US ARMY
AVIATION and
TROOP COMMAND

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035-1000

Test Techniques for Evaluating Flight Displays

LORAN A. HAWORTH AND RICHARD L. NEWMAN*
Ames Research Center

Summary

The rapid development of graphics technology allows for greater flexibility in aircraft displays, but display evaluation techniques have not kept pace. Historically, display evaluation has been based on subjective opinion and not on the actual aircraft/pilot performance. Existing electronic display specifications and evaluation techniques are reviewed. A display rating technique analogous to handling qualities ratings has been developed and is recommended for future evaluations. The choice of evaluation pilots is also discussed and the use of a limited number of trained evaluators is recommended over the use of a larger number of operational pilots.

1. Introduction

The head-up display (HUD) is becoming the primary fixed-wing flight reference for use during both visual and instrumental meteorological conditions. An offspring of the HUD technology, the helmet-mounted display (HMD), has been developed to accommodate the requirement for larger field-of-regard displays. The HMD is expected to become a primary rotary-wing flight reference in the future.

HUD and HMD allow the presentation of flight-critical information in a plethora of formats. This technology influx creates the potential for new and unique formats for information critical to flight and mission success to be conveyed to the flight crew. The historical methods of testing flight displays must be improved and updated to provide verifiable objective evaluations of HUD and HMD.

This document addresses the issue of evaluating the HUD or HMD symbology formats for use as primary flight references, although these observations apply to other flight displays.

1.1 A Brief History of HUDs and HMDs

The HUD is an outgrowth of World War II reflecting gunsights. Gunsights, which began as simple iron rings, developed into collimated displays reflected from a semitransparent combiner glass. The benefit of collimated virtual image for the pilot was the ability to focus on both the target and the sight, rather than having one appear blurred or doubled. The result of this development was the lead-compensating optical sight. Essential flight information in the HUD—such as airspeed or altitude—was also included to aid the pilot in maintaining an eyes-out orientation. As airborne computer graphics technology advanced over the next decades, the HUD evolved to a miniature instrument display.

The major advantages of HUD and HMD are seemingly obvious:

Reduced pilot workload– Pilot workload is reduced when the overall piloting tasks require head-up, outside-the-cockpit flight references.

Increased flight precision– The overlay of HUD/HMD-presented flight data on the external visual scene allows the pilot to fly more precisely.

Direct visualization of trajectory– A conformal display allows the pilot to directly assess the aircraft performance.

Increased flight safety– Essential flight information presented on the HUD/HMD reduces eyes-in the-cockpit during critical flight maneuvers.

In the early 1980s, an HMD was developed for the U.S. Army's AH-64 Apache attack helicopter. The AH-64 HMD, when integrated with the Pilot Night Vision Sensor system (PNVS), provided night vision information for pilotage and weapons aiming during nap-of-the-earth flight (ref. 1). Night vision video imagery from the AH-64 infrared sensor is combined with symbology for presentation at the HMD. While the PNVS system has increased the U.S. Army's rotorcraft night and all-weather operations capability during nap-of-the-earth operations, the added dimension of off-axis head movement and sensor video combined with symbology has added new challenges for symbolic displays.

1.2 Display Format Criteria

Since the late 1970s, a number of reports have been released citing significant deficiencies in HUD symbology and installation. The Air Force Instrument Flight Center

---

*Crew Systems, San Marcos, Texas.

(AFIFC) found HUDs were limited by serious drawbacks, including lack of failure detection, lack of standardization, and an increased tendency toward spatial disorientation (ref. 2). The HMD has only recently been introduced, so analyses and studies of these displays are not readily available.

While there are general specifications for military HUDs (ref. 3) and HMDs (ref. 4), the HUD symbology described has not been applied to any design. The helicopter HMD specifications agree with those of the AH-64.

Traditionally, electronic displays and the associated symbology have been procured as part of the airframe weapon system, not as part of "aircraft instruments." Classed as contractor furnished rather than government furnished equipment, adherence to general military standards and specifications has not been required for systems like the HUD. Symbology drive laws and dynamics are frequently missing from the specifications for both HUDs and HMDs.

Since HUDs were not considered "flight instruments," little need was seen to establish their suitability for use as a flight reference. Consequently, few flight procedures were developed and limited training was provided to pilots on how to use the HUD in routine flight.

The only HMD fielded to date (the AH-64) was principally introduced to enhance visual/forward-looking infrared (FLIR) cues for pilotage. As a result, AH-64 pilots are trained to use the HMD for flight purposes. However, the flight symbology has not been validated for use as a flight format. If a pilot enters instrument meteorological conditions (IMC) during low-level night flight, procedures dictate reverting to conventional panel-mounted instruments.

The reported deficiencies in both HUD and HMD would have been corrected during flight tests had they occurred in conventional panel instruments. However, because of an absence of performance based objective criteria, the HUD display evaluations have relied on subjective opinion polls.

1.3 The Future — Summary of Trends in Displays

Today's cockpit technology is progressing almost faster than we can write about it and advances in electronic display systems almost defy description. It seems certain that future transport and tactical aircraft will have cockpits with all-glass displays and, at most, a few conventional instruments for standby purposes. In addition, aerodynamic dictates of hypersonic transport or combat survivability may eliminate direct external vision in future cockpits.

HMDs will likely continue to progress from limited field of view (FOV) imagery presented to one eye to full FOV HMD presented to both eyes with improved resolution. To illustrate, by the late 1990s plans include progression from 30° x 40° FOV monocular HMD (in the AH-64) to a 30° x 52° total FOV dual optic HMD (in the RAH-66).

The basic question, however, remains: Will we develop a performance-based methodology for evaluation of HUDs and HMDs or continue to rely on a majority vote of pilots?

2. A Review of Display Symbology

2.1 Comparison of Displays

Table 1 lists some characteristics of traditional instruments and modern electronic displays. The conventional instrument panel (round dials) is characteristically fixed in position and has very limited ability to be programmed for different flight segments.

Conventional instruments can be color coded and are useful for displaying systems data. The pilot must look inside the cockpit to observe the instruments since they do not appear in the pilot's view of the real world.

Head-down displays (HDDs) using cathode ray tubes (CRTs) have many of the same characteristics as conventional panels, but it is possible to reprogram the same display for different phases of flight. For example, an electronic attitude (director) indicator can display different types and amounts of information during cruise, instrument approach, or takeoff. The electronic display can also generate symbology that is a real world representation, the contact analog. This has been extended to electronic moving map displays, which are analogs of the world when viewed from above. Finally, the electronic CRT display can integrate data from a number of sources, including the display of a velocity vector.

HUDs/HMDs share some of the characteristics of CRT displays. These are the abilities to be programmed, to time share, and to display integrated information from a variety of sources. Although color coding HUDs/HMDs has been discussed, it seems unlikely that either will have the same degree of color coding available in conventional or electronic head-down instruments anytime in the near future. Perhaps the most compelling difference between HUDs/HMDs and all other displays is the ability of the HUD/HMD to display real world conformal images.

Table 1. Display characteristics

Display characteristics	Round dials	HDDs[a]	PVDs[b]	HUDs	HMDs
In forward view			X	X	X
Collimated				X	X
Color coded	X	X			
Programmable		X		X	X
Time share		X		X	X
Integration possible		X		X	X
Foveal cues[c]	X	X		X	X
Peripheral cues[d]			X	*e*	*e*
Useful for systems	X	X		*f*	*f*
Contact analog possible		X		X	X
Conformal display possible				X	X
Can show flight path	X	X		X	X

[a] Head-down displays using CRTs.

[b] Peripheral vision displays.

[c] Foveal cues are those that require the pilot fix his attention on the display.

[d] Peripheral cues do not require the pilot's visual attention.

[e] Questionable value with restricted FOV.

[f] Caution/warning displays only. Additional system displays can add excessive clutter.

2.2 Published Specifications

A review of existing electronic display standards and specifications shows a limited number of standardization attempts. Current specifications and standards for electronic HUDs, HDDs, or HMDs are listed in table 2. Five of these specifications apply to military aircraft; four to civil transport aircraft; and one applies to both civil and military aircraft. Of the civil transport documents, two are industry recommended standards, one is an Advisory Circular, and one is a draft Advisory Circular.

HMD symbology standards are largely an outgrowth of existing standards for HUDs with the addition of specialized symbol and symbol driver requirements for hovering flight. To date, these specifications have had little impact on the development of any HUD or HMD.

There have been several critical reviews of HUD specifications. In the mid- to late 1970s, the U.S. Naval Aeromedical Research Laboratory reviewed existing HUD specifications and found a lack of data to substantiate these specifications (refs. 14 and 15).

In the mid-1970s, the Air Force Instrument Flight Center found that pilots had developed their own techniques for using the HUD and were, in fact, using the HUD as a flight reference (ref. 2). While the HUD did represent a significant aid as a flight reference, its reported usefulness was limited by several drawbacks: the lack of adequate failure detection, inadequate standardization, and a reported increase in tendency toward spatial disorientation. Follow-on studies have raised similar symbology issues.

In the early 1980s, two independent studies reviewed HUD specifications (refs. 16 and 17). These reviews found that there was little objective data to substantiate specifications, evaluations, or design choices. In the absence of objective performance data, most specifications were found to be based on subjective opinion. Furthermore, utility as a flight reference had not been considered.

Following these studies, the U.S. Air Force sponsored a program to develop HUD criteria. The result was a guide to assist the HUD designer to ensure that the next generation of HUDs would be adequate for their tasks (ref. 7). While providing design guidance, an evaluation methodology was still absent.

Table 2. Electronic display standards

Specification	Military	Civil	HDD	HUD	HMD	Reference
MIL-D-81641AS	X			X		3
MIL-STD-884C	X		X	X	[a]	5
MIL-STD-1295	X			[a]	X	4
MIL-STD-1787	X		X	X	[a]	6
AFWAL TR-87-3055	X	X		X		7
AFIFC TR-91-01	X			X		9
SAE ARP-4053		X		X		10
SAE AS-8034		X	X			11
FAA AC-25-11		X	X			12
FAA draft paper		X		X		13

[a] Not discussed in specification. However, the display type shown is within the scope of the specification.

In 1989, the U.S. Army Human Engineering Laboratory started a critical review of HMD requirements (ref. 8). Approximately 100 documents were reviewed and performance data were found to be lacking.

In the absence of objective requirements, all an evaluation pilot can do is determine the ability to fly by reference to the display without excessive workload. It is difficult to document an unacceptable display, particularly without performance criteria.

There have been a number of recurring problems with HUD specifications. The most common are:

No dynamic requirements– None of the government display specifications list any dynamic response requirements, other than "shall be free from unacceptable jitter." The specifications also fail to specify any sampling interval. As system capabilities grow, increased computer workload can force the computation interval to grow from 20–40 msec to 80–100 msec. At some point in the lengthening of this interval, the display quality will degrade dramatically.

There appears to be a misconception that 100 msec is a magic computation interval, below which there will be no display problems. This seems to be based on the idea of a 1/10 second human reaction time. In fact, sampling intervals of 100 msec can seriously degrade tracking in fighter aircraft (ref. 18).

Standardization– HUD specifications show a complete lack of standardization. As an example, in many HUDs the angle of attack (AOA) is shown by an error bracket that moves relative to the velocity vector. Some show a fast error, as the AOA bracket above the velocity vector, others reverse this.

The reason for choosing the fast error above/slow error below is based on the conventional fly-to philosophy common in navigation deviation indicators—if you are fast, pull the nose up.

The background for the reversed error sensing goes back to the Klopfstein format (ref. 19), which made use of the relationship between AOA, flight path angle, and pitch. The fly-from AOA error bracket was intended to emphasize this unique relationship.

With rational arguments favoring both the fly-to and fly-from senses, which is better in a HUD? At this point, the answer is not clear; however, it is obvious that having both arrangements in similar aircraft has the potential for problems. There should be an objective method of determining the better format. This method should be a performance-based evaluation.

Hidden specifications– Finally, there are several "hidden" specifications. One example was the 100 msec computer frame time mentioned before. Another is the precession that occurs as the airplane passes ±90° in pitch. This is a carryover from electro-mechanical attitude indicators to prevent gimbal lock. An electronic display has no need to keep this feature. Yet, many HUD designers feel that it is an essential feature—one designer even stated that there was a military specification requiring such a precession.

Gold-plated specifications– Many recent standardization attempts have been based on a "wish list" for HUDs that will do everything. In the civilian and military communities, the drafters of requirements assume that all future aircraft will carry wide FOV holographic HUDs with a complete inertial navigation system and precision distance measuring equipment (DME) available. The draft specifications appear to preclude non-conformal HUDs for many smaller corporate aircraft.

When drafting specifications and standards, there are places for displays with narrow FOVs driven by gyro platforms presenting air-mass data. These HUDs may not allow us to fly to Category III minimums, but they may still enhance the mission for which they are intended.

2.3 Need for Standardization

The need for absolute standardization in electronic displays is questioned. There appears to be a strong desire to have fighter HUD symbologies the same in all aircraft. This is surprising since there appears to be little or no standardization in fighter instrument panels.

The major reason for standardization is to reduce negative habit transfer and allow pilots to move rapidly from one airplane or system to another. Pilots today do not jump from one airplane to another and reaching back to prior training at critical points can be inappropriate.

In spite of this, standardization must play a secondary role to the effectiveness of the display for the particular aircraft and mission. While some aspects of standardization should not be changed arbitrarily (such as airspeed on the left and altitude on the right or the shape of some primary symbols), variations in mission, aircraft performance and agility, sensors available, and HUD FOV should allow flexibility in symbology standards. We should be surprised if a transport or a helicopter HUD were to look like a fighter HUD.

In addition, it is more important for modes within a given HUD to be consistent than to have standardization across aircraft. This argument is based on the pilot of a given aircraft who changes aircraft infrequently being exposed to multiple formats in the same aircraft on a daily basis.

For example, use of a variable compression pitch scale could have significant advantages during HUD instrument modes, but could present difficulties during an air-to-ground (A/G) weapon delivery mode. In this case, an A/G airplane should not use variable compression pitch scales in any mode, even if the "standard" instrument mode uses variable compression.

Historical HUD symbology problems were caused by inappropriate symbology, not by non-standard symbology. We must not become slaves to standardization for its own sake. Historical symbology standards may reflect the limitations of symbol generators at the time they were developed and should not be allowed to restrict development of advanced display formats. The primary goal should be enhanced pilot/aircraft performance *with HUDs designed and tested with mission performance in mind.*

2.4 Display Design Principles

Traditionally, display designers have sought expert pilot opinion for guidance during the development of new flight displays. While user opinion can be helpful, pilots tend to have diverse (and strongly held) opinions. In addition, pilots with limited background in display evaluation often limit the design of novel systems to those concepts with which they are familiar.

The display design must consider why the pilot needs the data and what the pilot is expected to do with the data. According to Singleton (ref. 20), the following questions should be considered during the display development:

1. Does the pilot's need justify the display?
2. Have all the necessary data been provided to the pilot? If not, what additional data are required?
3. Can the average pilot easily obtain the required data?
4. Does the display conform—
    - to the real world?
    - to other cockpit displays?
    - with previous pilot habits and skills?
    - with required decisions and actions?

Following completion of the display design, its evaluation should be based on objective, performance-based criteria and measures of the display's effect on mission performance. It is up to the evaluation team to determine what are suitable performance measures. These should reflect the intended mission of the aircraft and should include all mission segments.

All displays have a need to minimize display clutter and this is particularly critical with see-through displays. Since HUD/HMD symbols are presented in the pilot's view of the real world, obtrusive symbology should be kept to an absolute minimum. Not one "pixel" should be lit unless it "buys" its way onto the screen by providing a demonstrable improvement in performance (ref. 21).

3. Display Evaluation

3.1 History of Vote/Performance Evaluations

The following comments apply to evaluation methods, not to the particular displays or display concepts involved.

Performance based studies– In the 1960s, United Kingdom HUD studies were performance-based. Naish measured approach tracking performance and lateral and glideslope errors (ref. 22). One conclusion was that director symbols and slight pitch scale compression improved tracking performance. The shortcoming of the performance measures was the absence of measurements of the pilots' ability to maintain situational awareness in flight.

In the 1970s, Klopfstein developed a landing symbology as an aid to flying instrument landing system (ILS) approaches. This display featured a synthetic runway (a runway outline which appeared over the real runway) and used a unique angular presentation of AOA. Pilots who evaluated this display reported that precise airspeed control and tracking performance resulted even though no airspeed information was shown on the HUD (ref. 23). The conformal runway outline has been used in most civilian ILS HUDs (refs. 24 and 25).

In the mid-1980s, the U.S. Air Force studied the effect of HUD symbology on unusual attitude recovery and measured a variety of recovery parameters (ref. 26). The conclusions supported the early studies and recommended the use of compressed pitch scale and a recovery cue. This study also indicated that air-mass data might be beneficial. The conclusions lend weight to the need for an overall objective, performance-based test methodology.

In spite of these results, there has been reluctance to use the compressed pitch scale, the synthetic runway outline, or air-mass data in operational HUDs. This reluctance has not been based on performance-based evaluations, but on individual pilot opinions.

Opinion based decisions– The AOA bracket and the orientation of airspeed and altitude scales are two areas where conflicting opinion has created dissimilar formats to display the same information. The use of color coding for HDDs is another.

At one point, there were two quasi-standards for color HDDs developed by two competing transport airplane manufacturers. The HDD colors differed for scales and navigation symbols. On review, it appeared that once the decision to have the sky color be blue, the warning color be red, and so forth, had been made, only a limited number of choices remained. For example, if the sky is blue, the pitch scales cannot be blue also. Each company made a slightly different choice for various scales resulting in non-uniform colors.

A standardization meeting several years ago seriously proposed that a committee take an equal number of choices from each company's list and arrive at an "acceptable compromise." The alternative was to choose between the two companies. A performance-based evaluation was not discussed.

3.2 Subjective Data

Subjective pilot ratings play a key role in any display evaluation. Historically, pilot ratings have been patterned after one of two forms: the traditional Likert difficulty scale (ref. 27) or the Cooper–Harper Pilot Rating (CHPR) (ref. 28).

Likert rating scales– Traditional rating scales ask the pilot to rate the difficulty making choices as "very easy," "easy," "medium," "hard," or "very hard." A derivative of this type of scale is the task load index (TLX) rating scale developed by NASA (ref. 29). Similar ratings were used in previous HUD simulations (ref. 26). The chief advantage for a Likert scale is the ease with which a subject can learn them. It can also be useful for troubleshooting an unacceptable display.

One disadvantage of such scales is the reluctance of general subjects to use extreme values and the reluctance of pilot subjects to use "difficult" ratings unless the display is quite difficult to fly. As a result, a seven point scale frequently becomes a three point scale.

Cooper–Harper pilot ratings– The CHPR scale uses a decision tree to allow the pilot to "walkthrough" a series of dichotomous alternatives, by answering questions, such as "Is it [the airplane] controllable?"; "Is adequate performance attainable with a tolerable workload?"; and "Is it satisfactory without improvement?" Following these dichotomies, the pilot then makes a choice of three sub-alternatives.

The main advantage of this approach is that the logic tree involved produces consistent results—particularly with trained evaluators. This is evident in the area of aircraft handling qualities ratings.

A second advantage of the logic tree approach is apparent when evaluations are conducted without a control display or control symbology. In this case, we don't compare preferences, but determine if the performance objectives are met and what degree of pilot workload is required to meet them.

The major difficulty is the time that an novice evaluator must spend learning the logic tree. When using CHPRs with untrained evaluators, quite often a copy of the logic diagram is provided as an in-flight aid (ref. 18). Scales based on CHPR-type logic trees have been used during low altitude navigation targeting infrared for night (LANTIRN) evaluations (ref. 30). A similar scale, the Bedford workload scale, was used in the United Kingdom for HUD evaluations (personal communication with

J. Hall, Royal Aircraft Establishment, Bedford, England, 1990, and ref. 31).

It is imperative that a rating be taken in the context of a specific flight segment flown by a typical operational pilot. Cooper and Harper (ref. 28) emphasized this requirement, but it applies to all aircraft control-display evaluations as well. When using a task-oriented evaluation, the evaluator must use consistent performance standards. The standards should be related to operational standards, but must be clearly stated. Table 3 shows examples of such performance standards.

3.3 Display Evaluation

There are two aspects of flight displays that must be considered: can the pilot determine the value of a specific parameter, such as airspeed?; and can the display be used to control that variable? These two questions must be answered in the context of a specific task scenario. Because of the wide-spread acceptance of the CHPR scale in the flight test community, two logic trees were constructed to rate the readability and the controllability of displays (figs. 1 and 2). An earlier version of those figures (ref. 33) was used by the U.S. Army Center for Night Vision and Electro-Optics for Display Flight Assessment (ref. 34). The readability rating can also be applied to the ease of overall maintenance of situational awareness or attitude awareness.

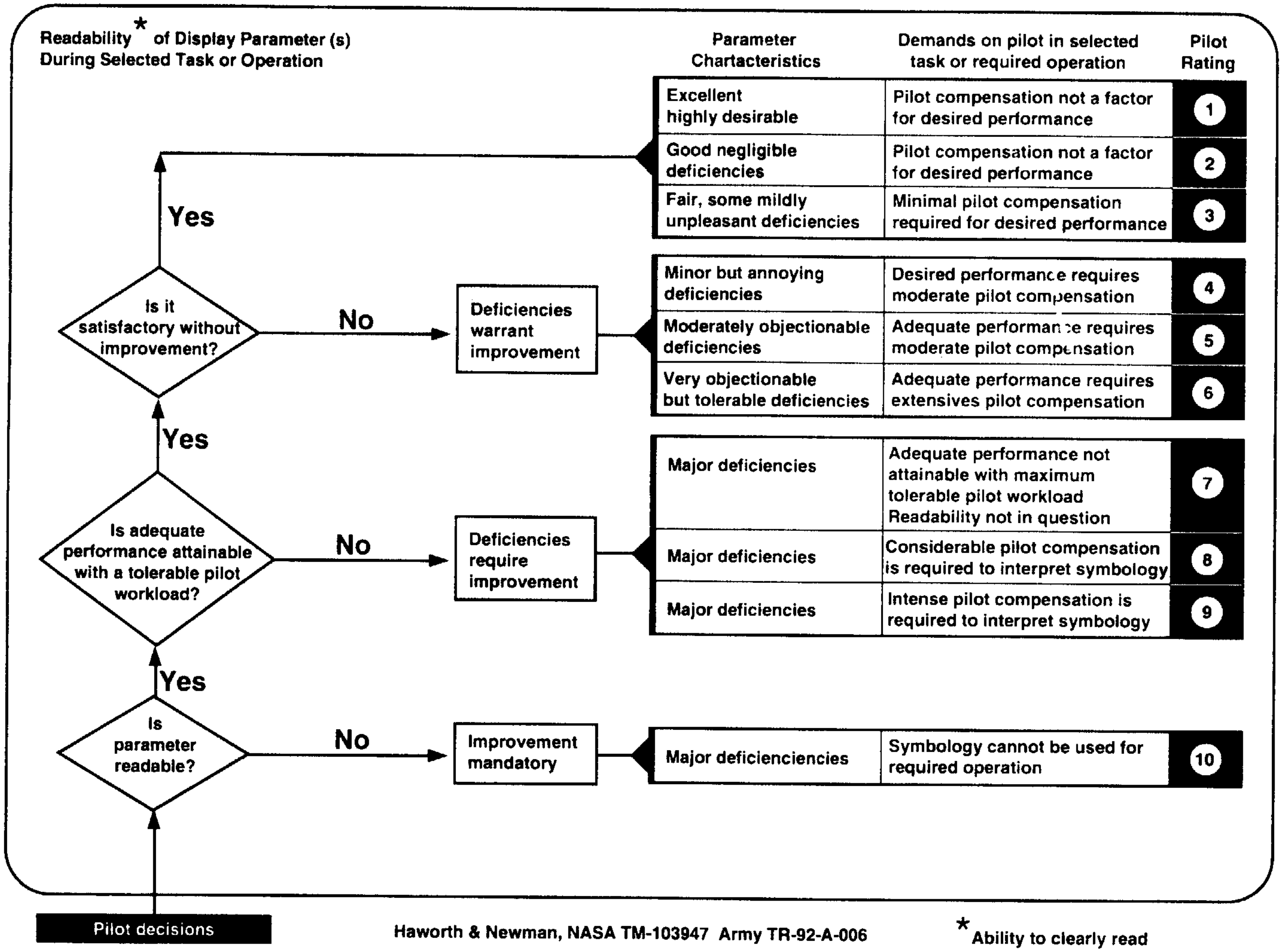
These display ratings follow the original Cooper–Harper decision tree closely. The difference between the display flyability rating and a handling qualities CHPR is the requirement that the evaluation pilot consider aircraft control using the display for information. This is essentially a CHPR of the airplane handling qualities in series with the display control laws. This rating for a given symbology will be expected to vary from aircraft to aircraft.

3.4 Additional Questions

In addition to the basic rating cards, questions should be asked addressing specific test issues, such as perceived problems with a particular display. These can be asked at the same time the rating card is completed (following each data run) or during a debriefing session. The final debriefing questionnaire should also ask for comparisons between the different displays.

Table 3. Evaluation task performance standards

Desired performance standards	Adequate performance standards
Dynamic maneuvers	
<2 sec to acquire new attitude. <5° heading and roll error at key points during maneuver. <3° heading error on recovery. <100 ft altitude loss.	<4 sec to acquire new attitude. <10° heading and roll error at key points during maneuver. <5° heading error on recovery. <200 ft altitude loss. No PIO.
Unusual attitude recoveries	
<1.4 sec to initial correct control input. Initial control input in accordance with published instrument standards (ref. 32). No control reversals. No overshoots on recovery to wings-level.	<1.8 sec to initial correct control input. Initial control input in accordance with published instrument standards (ref. 32). Single control reversal. Single overshoot on recovery to wings-level.
Instrument approach	
Loc/GS error <0.5 dot. Airspeed error <2 knots for 50% of task. No overshoots on intercept. Go around at DH +20/–0 ft.	Loc/GS error <1 dot. Airspeed error <5 knots for entire task. Single overshoot on intercept. Go around at DH +40/–0 ft

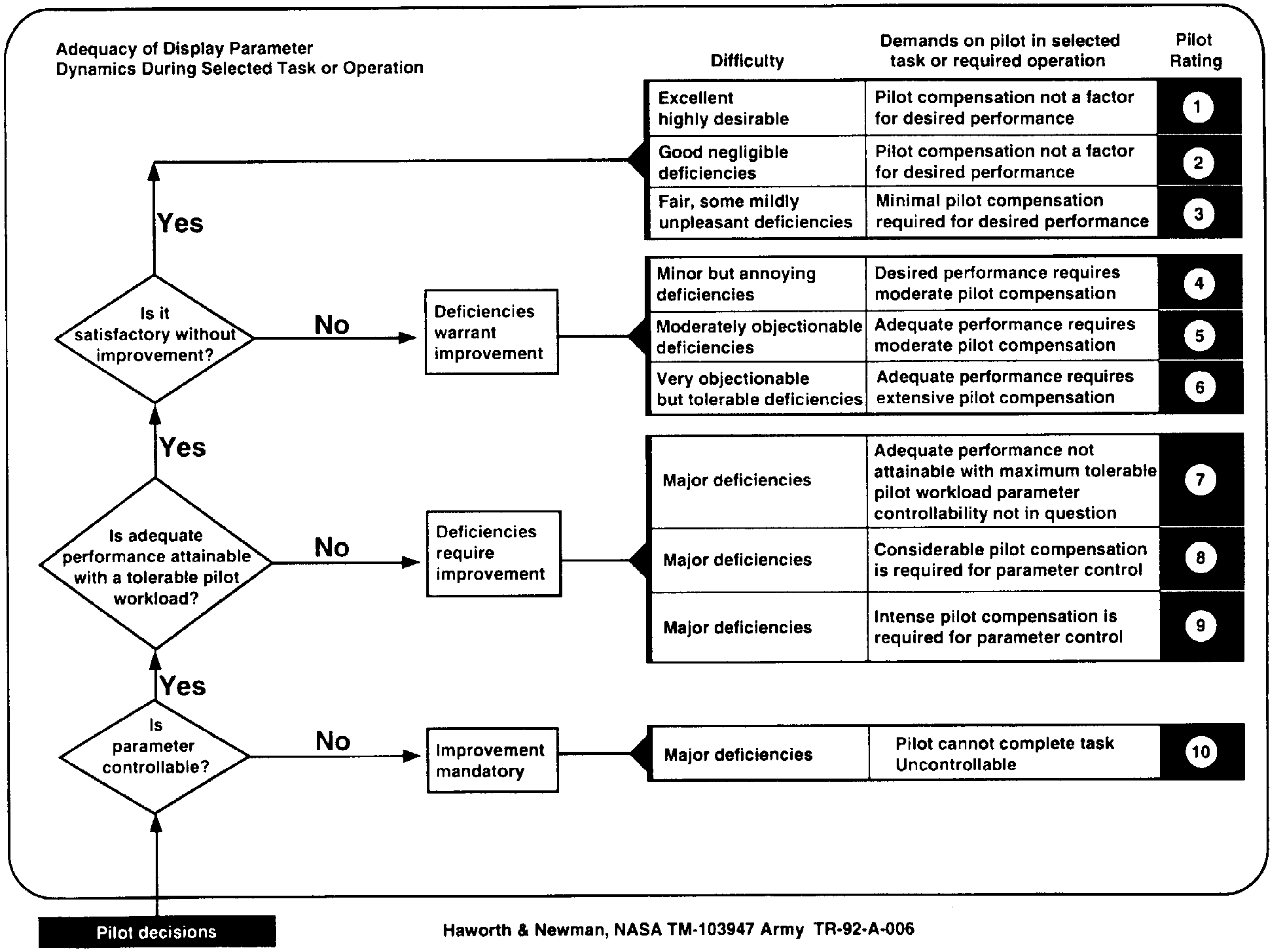


Parameter Charactesristics	Demands on pilot in selected task or required operation	Pilot Rating
Excellent highly desirable	Pilot compensation not a factor for desired performance	1
Good negligible deficiencies	Pilot compensation not a factor for desired performance	2
Fair, some mildly unpleasant deficiencies	Minimal pilot compensation required for desired performance	3
Minor but annoying deficiencies	Desired performance requires moderate pilot compensation	4
Moderately objectionable deficiencies	Adequate performance requires moderate pilot compensation	5
Very objectionable but tolerable deficiencies	Adequate performance requires extensives pilot compensation	6
Major deficiencies	Adequate performance not attainable with maximum tolerable pilot workload Readability not in question	7
Major deficiencies	Considerable pilot compensation is required to interpret symbology	8
Major deficiencies	Intense pilot compensation is required to interpret symbology	9
Major deficienciencies	Symbology cannot be used for required operation	10

Haworth & Newman, NASA TM-103947 Army TR-92-A-006

* Ability to clearly read and interpret parameter(s)

Figure 1. Display readability rating scale.



Haworth & Newman, NASA TM-103947 Army TR-92-A-006

Figure 2. Display controllability rating scale.

4. Evaluation Flight Tasks

All aircraft have many common mission segments: takeoff, climb, cruise, descent, terminal area maneuvering, approach to land, hover, and landing. For the most part, the problems during these common mission segments are universal. It has been said that most of our problems occur in the last 15 miles of the flight (ref. 35).

All mission tasks should be further divided to separate visual flight from instrumental flight. Each display has its particular set of problems.

When evaluating digital flight controls, the control system may be acceptable during routine mission tasks, but highly unacceptable during aggressive tracking tasks. This is described as a handling qualities "cliff." As the pilot tracks more and more aggressively, the handling qualities deteriorate quite suddenly and sharply, that is, falls off the cliff. This is often more pronounced during the landing flare or aerial refueling tasks (ref. 36).

Similarly, digital display dynamics can result in cliffs when evaluated during aggressive tracking tasks. For example, a velocity vector symbol may be well behaved until the pilot increases his gain to place it on a particular spot on the runway. For this reason, at least some of the experimental tasks should require aggressive tracking on the part of the subject pilots.

4.1 Evaluation Task Requirements

Evaluation tasks should be appropriate to the aircraft missions. Regardless of the mission, basic instrument and visual tasks must be flown, even if the display is intended for mission specific tasks only.

The tasks include aggressive pilot tracking to test the cliff. Low level terrain following, A/G tracking, landing flare, and unusual attitude recovery are examples of tasks requiring aggressive pilot inputs. For HUDs and HMDs, both instrumental and visual tasks should be flown.

It is also essential that dynamic maneuvering against real world backgrounds be flown, particularly when evaluating non-conformal pitch scaling or the effect of clutter. Flights against a real world background should be flown both day and night.

There must be some performance basis with which to compare different displays. Tracking accuracy is often used as a measure. Unusual attitude recovery uses reaction time to the first correct control input and the number of control reversals during the recovery.

4.2 Evaluation Tasks

The following tasks have been used in a variety of studies and are recommended as candidate evaluation tasks.

Unusual attitude recovery– This task involves a recovery from an unusual attitude using only HUD/HMD symbology. The airplane is placed in an unusual attitude and the subject pilot is directed to recover to a predetermined heading and altitude.

The head-down instruments are covered during this task and view of the real world cues blocked by the blue/amber system or another vision restriction device. During the entry into the unusual attitude, the HUD is blanked.

Additional unusual attitudes are introduced during other tasks, if possible. For example, during a simulated air-to-air tracking task, all external visual cues can be removed as though the target airplane flew into a cloud. The pilot has to recognize the situation and recover.

Dynamic maneuvering– This task involves aggressive instrument flight using only HUD/HMD symbology. The pilot is asked to fly a series of maneuvers appropriate to the airplane. Vertical S maneuvers modified to include abrupt changes of pitch and bank are suitable for this task. Instrument acrobatics (steep turns, barrel rolls, cloverleafs) are also used. At intervals, the subject pilot is distracted with a task requiring head-down viewing, such as reading a table arranged by rows and columns (personal communication with J. Hall, Royal Aircraft Establishment, Bedford, England, 1990).

Aimpoint tracking– Air-to-ground weapons delivery is a highly suitable experimental task for HUDs and HMDs. It requires aggressive tracking on the part of the subject pilot. For transports, a related task is a visual approach to landing requiring the pilot to maintain a specific aimpoint with the flight path marker.

This task helps to identify any problems associated with a non-conformal display.

Instrument approach– This task involves an approach to a landing or to a missed approach. Approximately half of the approaches are to a landing and half to a missed approach. Both precision and non-precision approaches are flown.

Visual approach– This task involves a visual approach to a landing. Some approaches are flown at night and both straight-in and circling approaches are flown.

System failures– During any of the tasks, it is important to consider the effect of system or sensor failures. ILS approaches should induce single axis failures (such as glideslope (GS) failure) and determine if the pilot can ecognize this event and maintain suitable performance following the failure.

4.3 Choice of Pilots

One fundamental question is: Should test pilots or operational pilots be used as evaluators?

Arguments favoring operational pilots include having pilots with recent mission experience. It is also possible to obtain a range of experience levels, from recent pilot training graduates to experienced pilots.

One problem with using operational pilots is that each pilot is often overtrained on a particular display and may be predisposed to that display—F-16 pilots prefer F-16 symbology, whereas F-18 pilots prefer F-18 symbology. Ideally, operational pilots with no symbology background should be used. Unfortunately, this is not possible. To avoid this problem, the experimenter must ensure that no particular symbology is overrepresented and that subjective data are used with care.

Another problem is the need to train operational pilots, both in how to fly with non-standard displays or techniques and in how to use rating scales. It is imperative that adequate familiarization and instructions are provided. This is most apparent with scales similar to the CHPR. The training can amount to two or three practice sorties per pilot compared with one for a trained evaluator.

Arguments favoring test pilots include having trained evaluators. Properly trained test pilots are used to rate airplane handling and should be familiar with rating scales such as the Cooper–Harper type of walk-through ratings. Test pilots are also skilled at communicating with engineers and can provide insight into display or control law problems.

Test pilots are experienced pilots, although, perhaps, not with recent mission experience. They usually have a broad range of experience in different airplanes and with different displays. This allows them to be able to adapt their individual control strategies to the display, such as using the pitch symbol versus velocity vector symbol for aircraft control.

The test pilot must remain objective. Special care must be taken if a test pilot has had a major role in designing the symbology. In this case, it would be best for the test pilot to be disqualified from the final approval portion of the tests.

The need to conduct practice sorties for untrained evaluators can quickly use up the available sorties in a program. For example, if 24 sorties are available, using two test pilots will allow for 22 data sorties. If six operational pilots are used instead, 12 to 18 practice sorties may be required allowing only six to 12 data flights.

If the display is novel or controversial, it may be necessary to use pilots with varying experience as a final check, although this will not normally be necessary.

5. Display Issues

5.1 Symbology

There are a number of symbology issues worth examining. However, space will limit the discussion to two current HUD issues.

One-to-one versus compressed scaling– For some time, it has been axiomatic in HUD designs that the display should be in one-to-one scaling with the outside world. While there is no doubt that such scaling is a considerable benefit to the pilot, there is also a growing amount of research indicating a benefit for compressed pitch scaling.

The main advantage of 1:1 scaling is that the pilot can immediately visualize his aircraft's trajectory if the HUD shows an inertial velocity vector. One-to-one scaling also allows for very precise determination of aircraft pitch attitude and immediate visualization of the aircraft's angle of attack (AOA) with an air mass system.

During ground-referenced maneuvers—A/G weapons delivery, low level navigation, approaches to landing, or obstruction critical takeoffs—visualization of the aircraft trajectory is critical. Using a scaled longitudinal acceleration to visualize aircraft trajectory, the pilot can determine the steady-state climb capability of his aircraft. Such a potential flight path can be beneficial during engine-out climbs, for example.

At the same time, early HUD research in the United Kingdom indicated that a pilot could fly a trajectory much more precisely if the pitch scaling were reduced to 1.5:1 or 2:1. This was in spite of not being able to detect smaller deviations as with 1:1 scaling (ref. 22).

Also, recent investigations into spatial disorientation indicates that compressed pitch scaling may help minimize the tendency to suffer spatial disorientation and may aid the pilot during unusual attitude recoveries (ref. 26). This same study suggested that a 1:1 HUD near the horizon combined with compressed scaling away from the horizon might be an acceptable compromise. Both continuously varying compression and a step change have been suggested. A step change is presently implemented in the F-16 HUD (personal communication with D. Howlings, GEC Avionics, Aug. 1991).

With these observations, there is a definite need to experiment to determine the effect of various HUD pitch scalings. The experiments must be performed in flight simulation and later in an airplane to validate or reject the simulator results. The experiment should explore the effect of variations in scaling, including both gradual and step changes. The effect of pitch scale compression (including variable gearing and step changes) should be evaluated during all ground-reference maneuvers.

Air mass versus inertial data– Recent research indicates that inertial quality attitude data are essential for HUDs. Designers have implicitly assumed that this requires the use of inertial velocity vectors as well.

The advantage of an inertial velocity vector is the direct display of the aircraft's trajectory against the real world. For example, when coupled with 1:1 scaling, the pilot can quickly determine exactly where the airplane is going, particularly during the final approach to landing.

At the same time, using an air mass velocity vector presents direct viewing of the aircraft's AOA. Air mass velocity vector, Klopfstein allowed pilots to fly more precise final approaches in terms of airspeed control and ILS tracking accuracy (ref. 19).

There is no question that pilots need to be aware of the aircraft performance in terms of the air mass. The issue is whether or not the benefits of displaying an air mass velocity vector is more important than the benefit of having a velocity vector conformal to the real world. If compressed pitch scaling becomes commonplace, the effect of a conformal velocity vector may be less apparent and may well influence the result. The issue must be evaluated with performance data and will certainly depend upon specific maneuvers and tasks.

5.2 Display Dynamics

In modern aircraft, the pilot obtains much of his flight information through the cockpit displays. It is not easy to separate display control laws from the aircraft handling qualities. The display dynamics, the seat-of-the-pants feel, and, during visual meteorological conditions (VMC), the view of the real world all form part of the feedback loop.

All of these feedback loops must be considered when performing evaluations. Traditional handling qualities evaluations only consider the aircraft dynamics with the motion and external vision feedback loops. Since these loops are essentially instantaneous, display dynamics do not affect the results.

Traditional instrument handling qualities evaluations used conventional instruments and benign instrument tasks. Because of the inherent damping in conventional instruments, and since typical instrument tasks are not very aggressive in nature, the instrument display dynamics do not interact with handling qualities demonstrations.

However, modern aircraft are being flown in aggressive maneuvers by reference to their displays. Pilots are dependent on the on-board sensors and associated displays. Even in VMC, the presence of an HUD or HMD ensures that the display dynamics cannot be ignored by the test pilot as they were in traditional VMC evaluations. The display has become an integral part of the aircraft and the display dynamics part of the overall control laws governing handling qualities.

The computer cycle time, or frame time, is an area of particular concern. Early avionics were electrical analogs of motion equations. Analog computers have the advantage of being much faster than digital computers and can process multiple functions in parallel. Digital computers, on the other hand, process multiple channels in series. The digital display computer has a defined cycle between 20 and 100 msec.

Data sampling will also adversely affect display dynamics. For example, if a given sensor input signal is sampled every 50 msec and this value is used to calculate the output signal that appears 50 msec later at the end of the cycle, two effects happen. First, the output is delayed 50 msec; second, the input and output are sampled every 50 msec. The waveforms of the signals are changed to reflect a series of step functions, not the smooth curves we expect. The sampling introduces high frequency "noise," which contaminates the input signal in addition to eliminating signal information at frequencies higher than the sampling frequency.

6. Conclusion

The rapid evolution of graphics technology allows greater flexibility in aircraft displays, but display evaluation techniques have not kept pace. Display evaluation in the past has been, to a large extent, based on subjective opinion and not on the actual aircraft/pilot performance. Pilot opinion, while valuable, must be tempered with performance measurement.

Modern digital displays interact strongly with aircraft dynamics and cannot be easily separated from the aircraft handling qualities. Many of the issues in fly-by-wire flight control systems are similar to flight displays issues. While technology allows the designer great opportunities to tailor the display to the mission, it also allows for greater opportunities for creating an unworkable system.

Display evaluation is not an simple task. It requires as much attention as any other flight critical system. Some of the problems high performance aircraft have exhibited in terms of lack of situational awareness or spatial disorientation have had their origins in poor display design. The community is concerned that informal discussions still form the basis for many symbology standards.

Future display evaluation methodology is not "cast in concrete" and is still in the developmental stage. The test and evaluation community should be challenged to participate in the discussions that are sure to follow, including the choice to use operational or test pilots, the choice of performance metrics and criteria, the choice of subjective rating scales, the need for standardization, as well as the many display issues themselves.

Display rating techniques analogous to handling-quality ratings presented in figures 1 and 2 were developed and are recommended for future evaluations. These display rating techniques were constructed to rate the readability, interpretability, and controllability of display symbology. An earlier version of these rating techniques was used to perform Integrated Helmet Display Flight Assessment by the U.S. Army Communication Electronics Command, Airborne Electronics Research Detachment (ref. 34).

It is important to recognize that while standardization is a desirable goal, there are good and valid reasons to deviate from a standard: new missions, new technology, new graphics processors, and different aircraft. We do not wish to constrain the design of new displays, but, rather, give the designer and the evaluator the display tools to allow them to make reasonable choices knowing the effect of their choices on the resulting performance.

References

1. Walker, D. J.; Verona, R. W.; and Brindle, J. H.: A Newly Developed Helmet Mounted Display for Attack Helicopters. Proceedings, Society for Information Display, 1980.
2. Barnette, J. F.: Role of Head-Up Display in Instrument Flight. AFIFC LR-76-2, 1976.
3. Military Specification, Display, Head-Up, General Specification for MIL-D-81641AS, 1972.
4. Military Standard, Human Factors Engineering Design Criteria for Helicopter Cockpit Electro-Optic Display Symbology. MIL-STD-1295, revision A in preparation, 1990.
5. Military Standard, Electronically or Optically Generated Displays for Aircraft Control or Combat Cue Information. MIL-STD-884C, 1988.
6. Military Standard, Aircraft Display Symbology. MIL-STD-1787, 1984.
7. Newman, R. L.: Head-Up Display Design Guide. AFWAL TR-87-3055, vol I, 1987.
8. Malkin, F. J.: AH-64 Helmet Mounted Display Lessons Learned. U.S. Army Human Engineering Laboratory Discussion Paper 40, 1987.
9. Bitton, F.; and Evans, R.: Report on Head-Up Display Symbology Standardization. AFIFC TR-91-01, November 1990.
10. Aerospace Recommended Practice: Flight Deck, Head-Up Displays. SAE ARP-4053, vol. 2, annex 8, 1986.
11. Aerospace Standard: Minimum Performance Standards for Airborne Multipurpose Electronic Displays. SAE AS-8034, 1982.
12. Transport Category Airplane Electronic Display Systems. FAA AC-25-11, 1987.
13. Criteria for Operations and Airworthiness Approval of Head-Up Displays for Air Carrier Aircraft. FAA Draft Advisory Circular, n.d.
14. Egan, D. E.; and Goodson, J. E.: Human Factors Engineering for Head-Up Displays: A Review of Military Specifications and Recommendations for Research. NAMRL Monograph-23, 1978.
15. Frank, L. H.: Comparison of Specifications for Head-Up Displays in the Navy A-4M, A-7E, AV-8A, and F-14A Aircraft. NAMRL SR-79-6, 1979.
16. Newman, R. L.; and Foxworth, T. G.: A Review of Head-Up Display Specifications. Crew Systems TR-84-04, Final Report on Contract F33615-83-C-5124, 1984.
17. Lovering, P. B.; and Andes, W. S.: Head-Up Display Symbology and Mechanization Study. Midwest Systems Research Report, Final Report on Contract F33615-83-C-5125, 1984.
18. Newman, R. L.; and Bailey, R. E.: Head-Up Display Dynamics Flight Tests. AFWAL TR-87-3055, vol. IV, 1987.
19. Klopfstein, G.: Rational Study of Aircraft Piloting. Intrados, ca. 1966.
20. Singleton, W. T.: Display Design: Principles and Procedures. Ergonomics, vol. 12, 1969, pp. 519–531.

21. Hughes, R. E.: The HUD Coloring Book. Naval Air Systems Command Unnumbered Document, 1991.

22. Naish, J. M.: System for Presenting Steering Information During Visual Flight: Part I, The Position of Presented Information, RAE TN-IAP-1132, 1961; Part II, The Form of the Presented Information, RAE TN-IAP-1138, 1962; Part III, The Influence of Errors and Limitations, RAE TR-64026, 1964.

23. Foxworth, T. G.: Flying the TC-121 All-Weather Head-Up Display. IFALPA Report, 1973.

24. Hamilton, F. W.: Development and Certification of a Commercial Head-Up Display. Proceedings 26th Society of Experimental Test Pilots Symposium, Beverly Hills, 1982, pp. 123–130.

25. System Description for FDI Model 1000 Head-Up Display. Flight Dynamics 404-0097, November 1983.

26. Newman, R. L.: Evaluation of Head-Up Displays to Enhance Unusual Attitude Recovery. AFWAL TR-87-3055, vol. II, 1987.

27. Murphy, G.; and Likert, R.: Public Opinion and the Individual. New York, Harper, 1938.

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

29. NASA Task Load Index (TLX), Version 1.0. Human Performance Research Group, NASA Ames Research Center, CA, 1988.

30. Papa, R. M.; and Stoliker, J. R.: Pilot Workload Assessment—A Flight Test Approach. In The Man-Machine Interface in Tactical Aircraft Design and Combat Automation, AGARD CP-425, 1987, paper 8.

31. Penwill, J. C.; and Hall, J. R.: A Comparative Evaluation of Two HUD Formats by All Four Nations to Determine the Preferred Pitch Ladder Design for EFA. RAE FM-WP(90)021, 1990.

32. Instrument Flying. U.S. Air Force AFM-51-37, July 1986.

33. Newman, R. L.; Haworth, L. A.; et al.: An approach for Display Evaluation. Presented at the 1991 European Society of Experimental test Pilots Symposium, Bath, England, May 1991.

34. Kimberly, J.; and Mueck, S.: Integrated Helment Display System (INVS) Flight Assessment. Airborne Electronics Research Detachment, Report NV-1-92, March 1991.

35. Hoerner, F. C.: HUD Military Aspects. Proceedings of the Flight Operations Symposium, Vancouver, 1979.

36. Berthe, C.; et al.: Fly-By-Wire Design Considerations. Proceedings 20th Annual European Symposium SETP, Linköping, 1988.

Abbreviations

A/G	Air to ground
AOA	Angle of attack
CHPR	Cooper–Harper Pilot Rating
CRT	Cathode ray tube
DH	Decision height
DME	Distance measuring equipment
FLIR	Forward-looking infrared
FOV	Field of view
GS	Glideslope
HDD	Head-down display
HMD	Helmet-mounted display
HUD	Head-up display
ILS	Instrument landing system
IMC	Instrument meteorological conditions
LANTIRN	Low Altitude Navigation and Targeting Infrared for Night
PIO	Pilot-induced oscillation
PNVS	Pilot Night Vision Sensor
TLX	Task Load Index
VMC	Visual meteorological conditions

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY ***(Leave blank)***

2. REPORT DATE
February 1992

3. REPORT TYPE AND DATES COVERED
Technical Memorandum

4. TITLE AND SUBTITLE
Test Techniques for Evaluating Flight Displays

5. FUNDING NUMBERS
505-64-36

6. AUTHOR(S)
Loran A. Haworth and Richard L. Newman

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Aeroflightdynamics Directorate, U.S. Army Aviation and Troop
Command, Ames Research Center, Moffett Field, CA 94035-1099 and
Ames Research Center, Moffett Field, CA 94035-1000

8. PERFORMING ORGANIZATION REPORT NUMBER
A-92139

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration,
Washington, DC 20546-0001 and
U.S. Army Aviation and Troop Command, St. Louis, MO 63120-1798

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
NASA TM-103947
USAATCOM TR-92-A-006

11. SUPPLEMENTARY NOTES
Point of Contact: Loran A. Haworth, Ames Research Center, MS 243-4, Moffett Field, CA 94035-1000
(415) 604-6944

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified — Unlimited
Subject Category 05

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 words)***

The rapid development of graphics technology allows for greater flexibility in aircraft displays, but display evaluation techniques have not kept pace. Historically, display evaluation has been based on subjective opinion and not on the actual aircraft/pilot performance. Existing electronic display specifications and evaluation techniques are reviewed. A display rating technique analogous to handling qualities ratings has been developed and is recommended for future evaluations. The choice of evaluation pilots is also discussed and the use of a limited number of trained evaluators is recommended over the use of a larger number of operational pilots.

14. SUBJECT TERMS
Flight testing, Flight symbology, Helmet-mounted display, HMD, Head-up display, HUD, Flight displays, Electronic display standards, Display evaluation, Display readability rating scale, Display controllability rating scale

15. NUMBER OF PAGES
18

16. PRICE CODE
A02

17. SECURITY CLASSIFICATION OF REPORT
Unclassified

18. SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

19. SECURITY CLASSIFICATION OF ABSTRACT

20. LIMITATION OF ABSTRACT

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102