Modified Cooper Harper Evaluation Tool for Unmanned Vehicle Displays

M. L. Cummings, Kevin Myers, Stacey D. Scott

***Abstract*—Operator interfaces are typically the sole information conduits between unmanned vehicles (UVs) and their human operators. Thus, it is essential that information be presented to operators clearly and efficiently. However, there is no current standardized methodology by which to subjectively evaluate unmanned vehicle displays to ensure they provide operators with sufficient information for effective mission performance. As a result, it is challenging for system developers to assess their display designs. To address this issue, this paper presents a new evaluation tool for subjectively assessing unmanned vehicle displays. Our evaluation tool, the Modified Cooper Harper for Unmanned Vehicles Displays (MCH-UVD), modifies the commonly used Cooper-Harper manned aircraft assessment tool by shifting emphasis away from evaluating the physical control of an aircraft, to evaluating how well the displays support basic operator information processing. The MCH-UVD tool provides a standardized rating scale to help determine whether a UV display enables the information gathering and processing necessary to complete and manage higher-level system tasks.**

***Index Terms*—technology assessment, unmanned vehicles, user interface human factors, user interfaces**

I. Introduction

UNMANNED vehicle (UV) systems currently, and for the foreseeable future, require considerable input and interaction with a human operator. Thus, an effective operator interface is critical to the success of UV operations. However, determining what constitutes effective interface design in such complex systems can be quite challenging. Typical interface design and usability testing methods are not designed to evaluate interfaces for their ability to support information analysis and problem-solving, especially in situations where there is high levels of information uncertainty and limited time to process that information. This lack of interface assessment techniques also makes it difficult to make comparisons across a set of candidate operator interface designs in existing or proposed UV operator interfaces.

Unlike productivity losses that can result from poor user interface design in business applications, the consequences of providing an operator a poorly designed UV system interface can be dire. With the increasing involvement of unmanned vehicles in military and homeland security operations, search and rescue, and disaster relief efforts, ineffective UV operator interfaces could result in human casualties. Such disastrous consequences could be caused by confusion and errors as a result of an operator's inability to obtain, interpret, or act on critical sensor information.

One possible way to address the need for a standard evaluation tool for UV displays is to draw on measurement techniques from a similar domain. The aviation industry has long relied on a standardized measurement tool called the Cooper-Harper rating scale to assess the controllability of manned aircraft. A significant benefit of the Cooper-Harper evaluation tool is that it provides a standardized scale that can be used to compare the handling qualities across aircraft. An additional benefit of the Cooper-Harper evaluation tool is that it helps test pilots articulate specific types of aircraft handling problems, which can then be addressed in any re-design efforts on the aircraft under evaluation.

However, unlike in manned aircraft operations, the physical control of the vehicle is typically not the primary task for UV operators, especially as UV automation increases. As the operator role shifts away from teleoperation towards supervisory control of mission tasks, an assessment tool is needed that addresses higher-level cognition instead of lower-level control skills. In order to create a more suitable measurement tool for assessing the effectiveness of a UV display given this higher-order cognitive tasking, we modified the original Cooper-Harper rating scale to focus more on the ability of the UV system to support the operator's role as information analyst and decision-maker.

Before presenting our proposed rating tool, called the Modified Cooper-Harper for Unmanned Vehicle Displays (MCH-UVD), we first provide background information on the original Cooper-Harper rating scale. The paper then details the MCH-UVD evaluation tool, followed by a discussion of some preliminary experiences in using this tool to assess existing and proposed UV displays.

M. L. Cummings is with the Humans and Automation Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (corresponding author: 617-252-1512; e-mail: missyc@mit.edu).

Kevin Myers is with the Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: kmmyers@mit.edu).

Stacey D. Scott is with the Humans and Automation Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: sdscott@mit.edu).

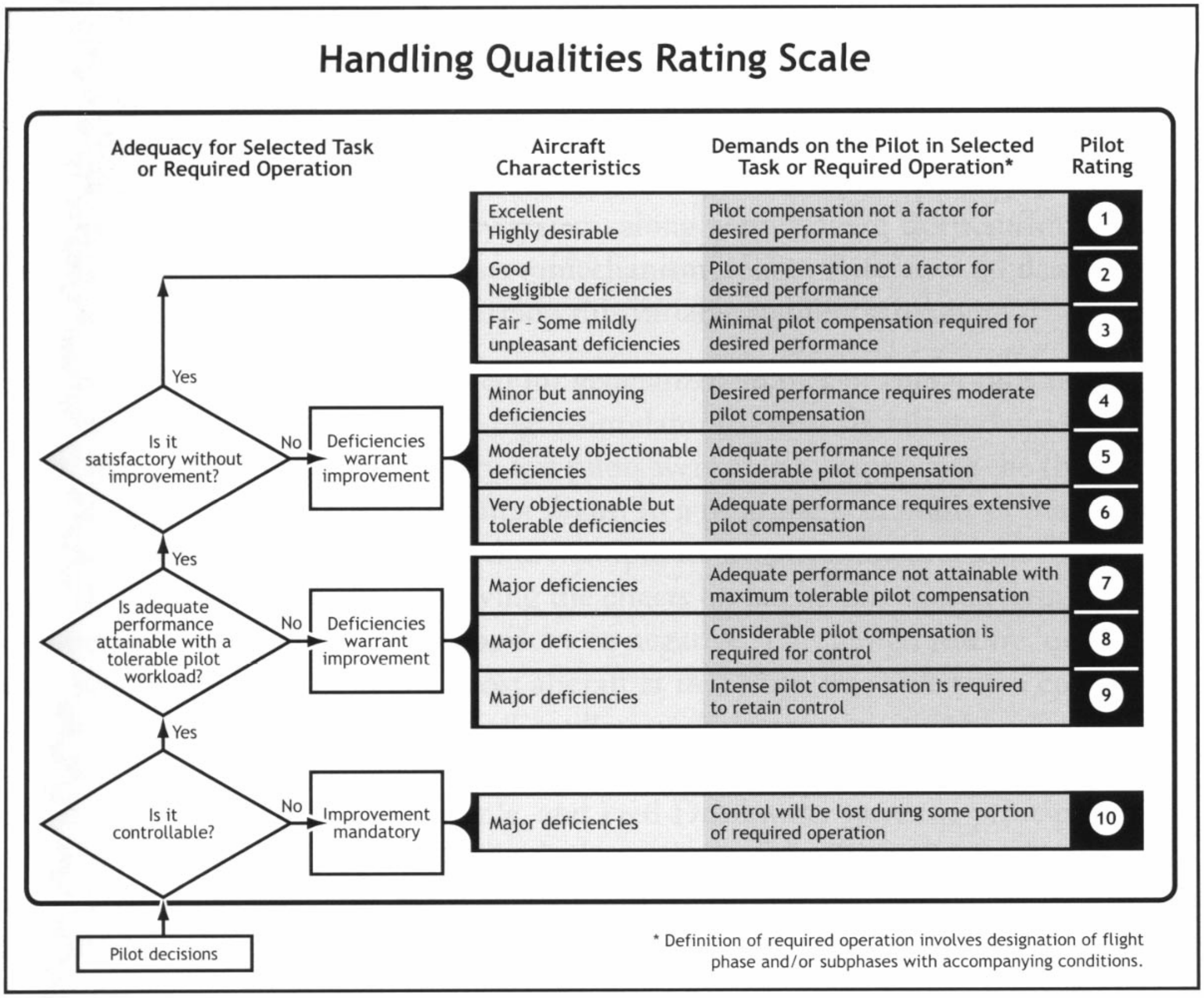

Fig. 1. Cooper-Harper Scale.

II. The Cooper-Harper Scale

In 1957, NASA scientist George Cooper presented a rating scheme that attempted to quantify how a pilot's workload affected task performance. This evaluation became known as the Cooper Pilot Opinion Rating Scale and was soon applied over years of flight and simulation testing. Later, through the assistance of Robert Harper, the original scale was modified to better assess the handling characteristics of an aircraft. This new evaluation was renamed the Cooper-Harper Handling Qualities Rating Scheme [1] (see Fig. 1), and remains today as an enduring subjective measure of aircraft design and performance [2].

As is seen in Fig. 1, the pilot evaluates the airplane's flight handling based upon controllability, workload, and attainable performance goals. The pilot then quantifies his opinion by describing the aircraft controllability characteristics and selecting the demands he experienced in flight while performing certain tasks. These demands then map clearly to a 1-10 scale which can then be used for statistical analyses. Thus the Cooper-Harper is a quasi-subjective rating in that it translates a subjective assessment to a quantitative and continuous metric. This rating assists engineers by describing which design features are sufficient and which are in need of improvement. The original Cooper-Harper Scale served as a valuable tool in assessing airplane handling characteristics since its inception.

III. Modifying the Cooper-Harper Scale

The key distinction for the adaptation of the original Cooper-Harper is that physical control of a UV is not a primary task for the operator, which is especially true for UAVs (unmanned aerial vehicles). While the control is an essential component of the mission, a UAV operator is more concerned with

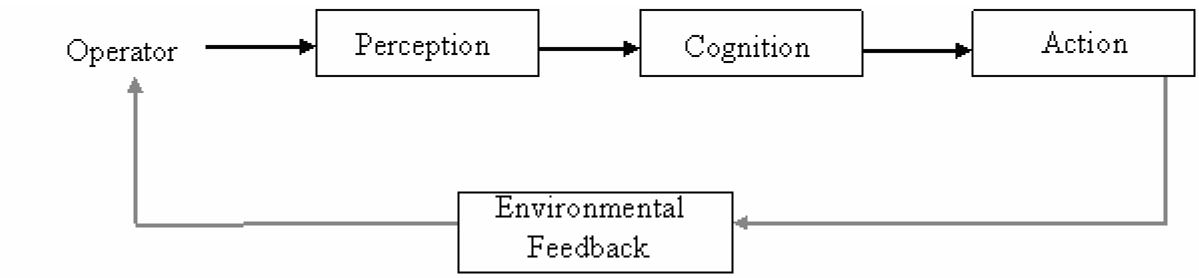

Fig. 2. Information Processing for Display.

higher-level system tasks since automation essentially controls, at a local level, the aircraft. Indeed, there is no true manual control of a UAV because even in the "manual" mode, the control is still fly-by-wire and mediated by a computer. Thus, in order to apply this scale to UAV control, a modification of the traditional Cooper-Harper scale was required to shift emphasis away from evaluating the physical control of an aircraft, to evaluating how well operators can achieve higher level goal tasking. As such, a subjective rating scale must provide an cognitive analysis of a remote operator's ability to effectively and efficiently complete and manage higher-level system tasks.

Moreover, with the inundation of unmanned vehicle ground control stations (or laptops, PDAs, etc.) that provide operator control, there is no current standardized methodology by which to subjectively rate the cognitive "handling qualities" of displays that are the critical window to safe and efficient vehicle operation. Just as the original Cooper-Harper scale allows pilots from many different companies and agencies to rate many different aircraft for direct comparison, our proposed Modified Cooper Harper for Unmanned Vehicles Displays (MCH-UVD) will allow original equipment manufacturers, government agencies, and customers of unmanned vehicle technologies the ability to draw comparisons of competing displays as well as understand how, where, and why a single display is either supporting or degrading operator cognitive processes.

IV. Construction of the MCH-UVD

Since a display is the primary communication link between unmanned vehicles and humans, it becomes a critical information conduit. Thus, it is essential that information be presented to the operator clearly and efficiently. To accomplish this, the display must accomplish three things. First, the display setup must allow information to be easily acquired. Second, the acquired information must be presented in a way that lends itself to efficient analysis. Finally, the display must assist the operator in decision-making. These three steps follow the general information processing model expressed in Fig. 2.

A modified Cooper-Harper Scale which essentially captures those basic information processing elements for the evaluation of UV displays can be found in Fig. 3 The idea is similar to the Cooper Harper in that an operator enters the flowchart from the bottom and decides whether or not the conditions apply, navigating through the decision points to finally arrive at a quasi-subjective rating of the system. As discussed previously, this is quasi-subjective in that for any given display and task, an operator will express a subjective opinion. However, there is an objective component to this display as each of the final 10 ratings has specific conditions that must be met. Thus even though the ratings are somewhat subjective, objective parameters can be mapped explicitly to the opinion and the numerical rating then provides a conduit for objective statistical analysis.

The ten ratings, which represent increasingly difficult human-vehicle display interaction, are discussed in detail in the next sections. They represent the information processing flow illustrated in Fig. 2, in that the first ratings address perceptual display components, followed by those that support analysis processes, and then finally those display elements that support decision-making and action.

A. Perception (Information Acquisition)

Major Deficiency–Poor Design: Rating 10

When designing a display, one of the most basic requirements is that operators can find the information they need. A display that does not accomplish this can be classified as a poor design. Under this classification, a display can fail in three ways. The first of these is through the omission of essential data that is required to manage a mission. An example of this could include failing to display the UAV's altitude, speed, or current location. The next basic failure would be to place critical information in a location that makes acquisition impossible or unrecognizable. This could occur when critical information is compiled in a large list of values or buried deep with the display architecture. This can also lead to the third and final failure which occurs when information retrieval time is so excessive that the information is irrelevant upon discovery. Any occurrence of these failures will assign the display a rating of ten which would require a mandatory redesign of the display. Without correction of the errors identified under this rating, the mission will fail due to the operator's inability to perceive critical information and ultimately manage the mission.

Major Deficiency–Attention Interference: Rating 9

This level involves two types of problems with respect to the operator's attention. The first problem occurs when a

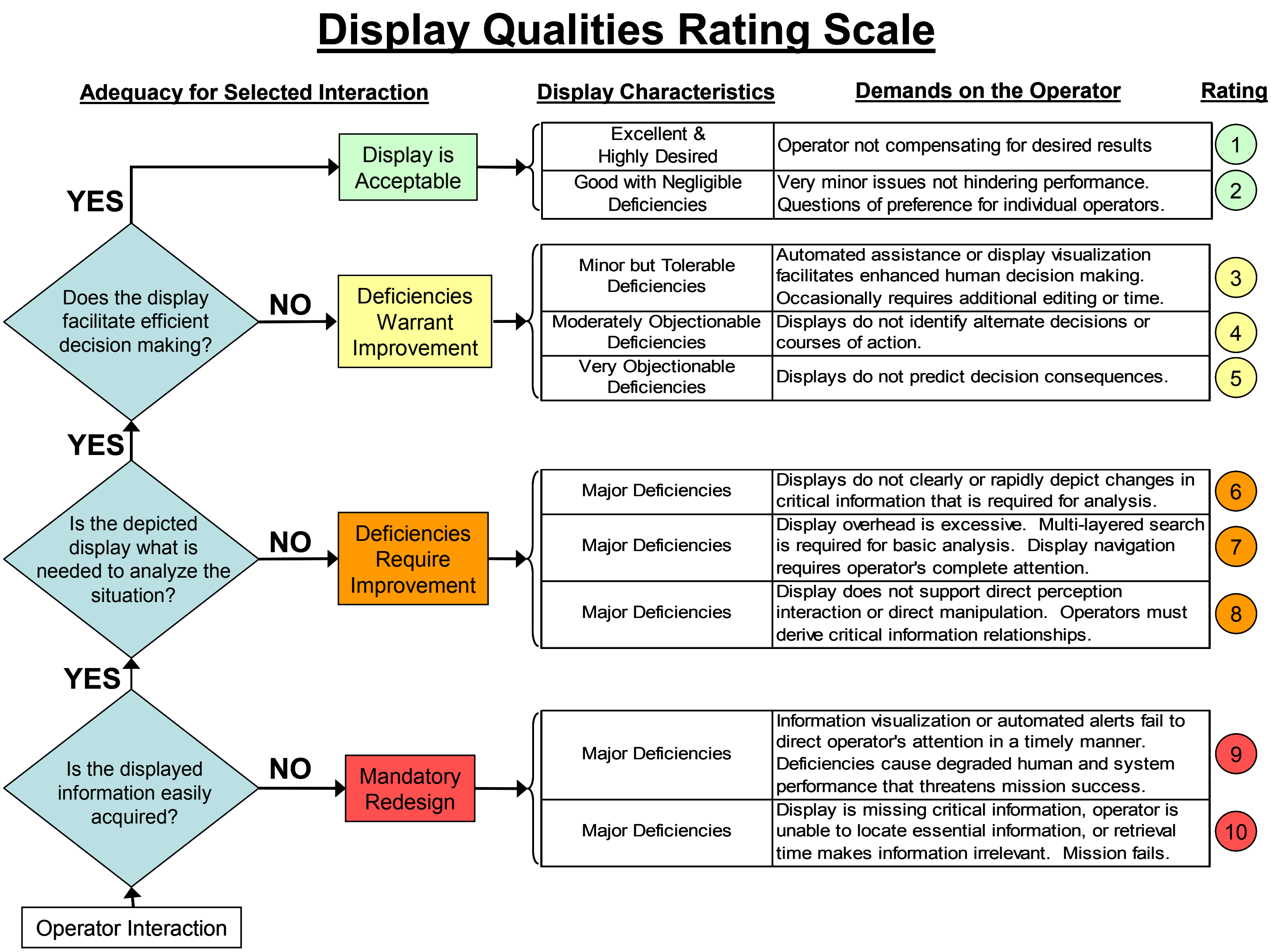

Fig. 3: Modified Cooper-Harper (MCH-UVD) Scale.

design's secondary characteristic interferes with a display's primary purpose, resulting in distraction. An example of this could be a display that alerts the operator by creating pop-up windows which block the operator's view of the desired or mission critical information. In addition, the cumulative effort of continually having to close them could also distract the operator from other tasks. The second type of problem under this classification is a failure to alert the operator to some abnormal state. If a display fails to direct an operator's attention towards a mission critical event, there is a significant flaw in the display design. An example of this could include a failure to notify the operator that the UAV must abort its mission due to fuel endurance. These two types of problems constitute a major deficiency and require a mandatory redesign of the display. Without correction of the problems identified under this rating, the operator's performance will be severely degraded and the success of the mission will be threatened.

B. Cognition (Information Analysis)

Major Deficiency–Direct Perception: Rating 8

With the required information at hand, the operator is ready to begin problem solving by analyzing the situation. However, under the lowest rating of information analysis, the operator must determine what information sources are needed for the analytical tasks. If operators must search for disparate sets of information (either across displays or within a set of windows within a display), and then mentally integrate this information, the cost is high in terms of mental workload, likelihood of errors, and time for analysis. Classification of this deficiency results in a rating of eight on the modified scale and requires improvement.

Displays should not force the operator to derive critical information requirements that could directly impact the outcome of the mission. Instead, displays should support direct perceptual interaction or direct manipulation, which allows users to employ more efficient perception processes rather than the cognitively demanding processes involved when relying on memory, integration, and inference [3]. An example of this is the determination of safe operating limits for the engine. Operators should not have to remember what the upper or lower limits are, for example, for engine temperature. This information should be clearly displayed, preferably in graphical format so operators can immediately perceive how far out of limits a setting is to determine criticality.

Major Deficiency–Cognitive Resource Allocation: Rating 7
While all information cannot readily be displayed on a single screen, the ease at which an operator navigates through the display architecture must be considered. Displays must be designed with the understanding of which screens will be commonly utilized together, the frequency with which a page will be accessed, and how many sub-levels will be required to find information. Displays that earn a rating of seven fail to address these considerations in their design. Displays with excessive overhead require the operator to utilize cognitive resources when they are not needed. Operators are forced to access their long-term stores on how to navigate through the display instead of concentrating on analyzing the problem. This increase in mental workload elevates the likelihood of error and could prove disastrous for complex analytical tasks. Thus, the poor allocation of cognitive resources demanded by this deficiency requires improvement.

Major Deficiency–Change Blindness: Rating 6
A final concern related to information analysis involves interface problems that contribute to operators missing critical or changing information. An operator cannot correctly solve problems if he/she does not have the most current information. This can occur when an operator is concentrating intently on one task and fails to notice a change in the status of the UAV. Originating from a failure to clearly or rapidly depict critical information changes, such events arise from cluttered display screens or poor change discriminators. For example, if an operator fails to notice that the UAV is losing fuel from battle damage because he is plotting a new route around a threat, a change blindness failure has occurred. As such, the outcome of the mission is questionable, and a design improvement is required to fix this discrepancy.

C. Action (Decision-making)

Very Objectionable Deficiencies–Uncertainty: Rating 5
At this point in the operator's interaction with the display, we are concerned with the issue of efficient decision-making. Specifically, rating level five examines the effects of the operator's uncertainty and the consequences of his decisions. An ideal display will attempt to predict the future results of a choice by displaying possible constraint violations (a proposed plan that will cause the UAV to run out of fuel) or predicting the probability of a certain outcome (a proposed plan that will likely put a UAV in a threat area) . Displays should attempt to negate an operator's inadequate understanding of the situation, lack of information, or conflicted view of alternatives. For example, by displaying how the observation of an additional target will affect UAV endurance, an operator will have improved decision-making abilities. Ultimately, this capability can only enhance the possibility of a successful mission. As such, a display that lacks this ability has a deficiency that warrants improvement.

Moderately Objectionable Deficiencies–Risk: Rating 4
A display at this level of deficiency fails to provide the operator an ability to examine multiple solutions or alternatives to a decision. While this process probably involves a computational heuristic or algorithm, a display that is capable of displaying a variety of alternatives would of great value to the operator. An example of this would be a display that prioritizes targets for observation based upon prioritization and threat probability. A display that lacks such support for option analysis and risk assessment has a deficiency that warrants improvement.

Minor but Tolerable Deficiencies–Automated Assistance: Rating 3
A display that provides some level of automated assistance of display visualization can facilitate enhanced operator decision-making. However, if the automated solution or display occasionally requires additional editing for improvements, time may have been more wisely utilized by a purely human-derived solution. For example, automated path planners that require operator intervention can often cause a higher workload than had the operator not used the automation at all. Moreover, operators should be allowed to set contextual automated alerts in the system for tasks management like "Alert me when a UAV comes within ½ mile of a no-fly area." These could be set either by individual operators or conform to unit policy. A display that uses unreliable automated planning tools or does not provide tools for contextual filtering, sorting, and alerting has minor deficiencies that warrant improvement.

D. Acceptable Displays

Good with Negligible Deficiencies–Minor Issues of Preference: Rating 2
At this point in the rating scale, the display has met an operator's requirements for information acquisition, information analysis, and decision-making. At this level, the operator only will only have very minor issues of preference that do not affect performance. Examples of this type of deficiency are non-preferred font size, font type, or display color. No changes to design are required.

Excellent and Highly Desired–No Issues: Rating 1
In this rating, the operator is not compensating for any desired results. He/she is completely satisfied with the display and would alter no properties.

E. *Examples of Display Qualities Rating Scale*

For each of the 10 levels previously described, Table 1 gives examples of what deficiencies would cause an operator to select that particular level.

Table 1: Example Ratings for Level Selection

Rating	Sample Interface Deficiencies
1	No compensation required.
2	Non-preferred window layout
3	Automation's recommended route routinely requires editing to meet operator's objectives.
4	Prioritized targets not depicted.
5	The impact of a new target on UAV endurance or return-to-base time is not displayed
6	Operator's focus on other tasks prevents recognition of change. For example, operator fails to notice UAV rapidly loosing fuel because he/she is plotting new route around the threat.
7	Operator must concentrate completely on display navigation.
8	Operator cannot simultaneously identify targets and monitor UAV altitude and airspeed.
9	Cluttered screens or overlapping windows.
10	Display does not include fuel endurance, altitude, aircraft location, etc.

V. MCH-UVD Rating Scale Case Studies

In order to determine the MCH-UVD rating scale effectiveness as an evaluation tool for UV operator interfaces, we have used the MCH-UVD to gather subjective assessments of UV operator interfaces in two separate user studies.

A. *Case Study 1*

The first study involved a partial evaluation of an older version of a VCS operator interface for UAVs like the Shadow 200. During this study, five subjects (all college students with no UAV experience) were first given basic training with the VCS system and then asked to perform three tasks involving the VCS vehicle control and warnings display screens (Fig. 4). After completing each task, subjects then used the MCH-UVD to rate how well they felt the VCS operator interface supported the activities of that particular task. In addition to providing a numeric rating, subjects were also asked to describe their rationale for choosing the given rating.

The study tasks involved completing pre-flight inspection checks and mission preparation (preflight task), performing a standard procedure for transitioning from friendly airspace into enemy territory (enemy territory task), and performing some typical actions related to diagnosing and completing emergency actions during the failure of one of the UAV systems (malfunction task).

The results of the study reveal that subjects generally felt that the operator interface provided poor support for the preflight task. As shown in Fig. 5, four of the five subjects submitted MCH-UVD ratings of 9 or 10 (Mean (M)=8.0, Standard Deviation (SD)=2.8, Median (MD)=9), indicating that they felt the VCS operator interface contains major deficiencies for supporting preflight activities, warranting redesign. Subject comments support this conclusion, with typical remarks such as: "[the] information acquisition is not intuitive" and "critical information is missing."

The assessment of the operator interface for the enemy territory task produced somewhat better ratings. Most subjects felt that the interface supports basic information acquisition for performing the entering enemy territory task, which involved altering various vehicle flight, sensor, and communication settings. However, the MCH-UVD ratings (M=7.8, SD=1.9, MD=8) suggest the operator interface could be improved to better support this task, confirmed by subject comments such as "the display needs a better tool for adjusting altitude and airspeed."

The VCS operator interface received the best MCH-UVD average ratings for the malfunction task, which involved the subjects determining the type of malfunction their vehicle was experiencing, and then performing a number of emergency

(a)

(b)

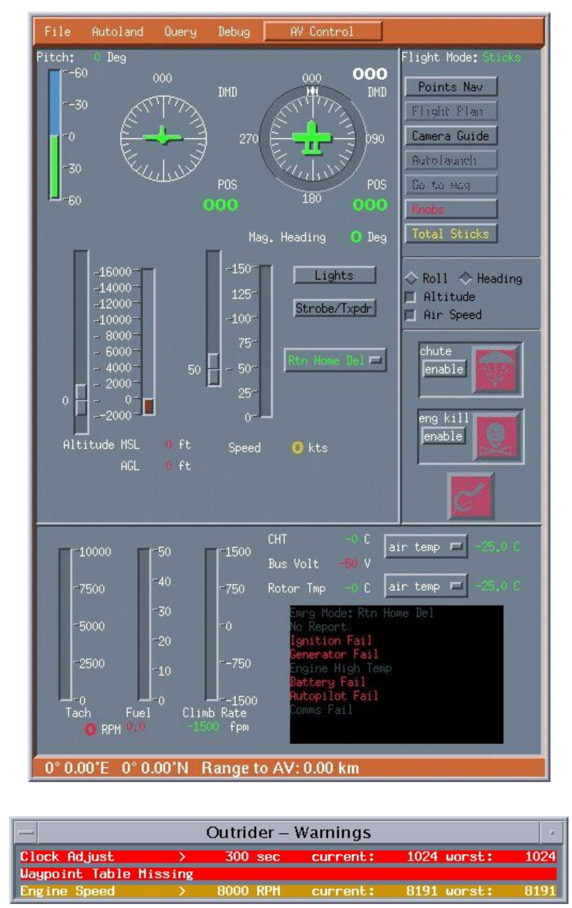

Fig. 4. Example VCS vehicle control (a) and warnings (b) displays [4].

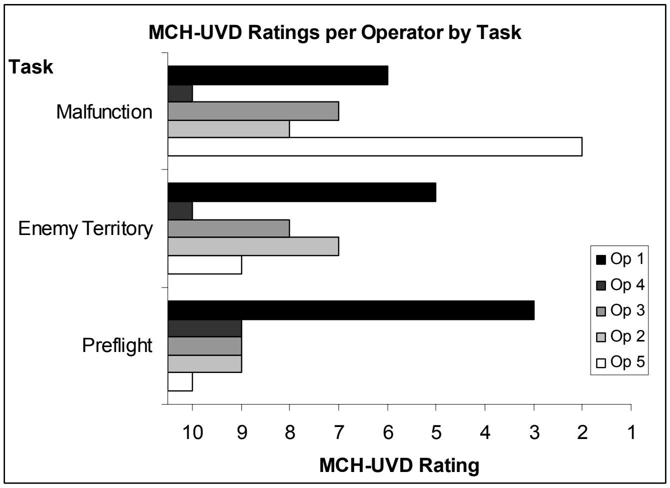

Fig. 5. Reported MCH-UVD ratings of the VCS operator interface per operator for the three study tasks.

procedures to safeguard the UAV. This task however, produced the largest variation in subjective assessments, with MCH-UVD ratings ranging from 2 to 10, with the median score = 6, indicating some improvements were warranted to support this task. Some reported comments indicate frustration in that "critical information is buried under various tabs." However, it appears that not all subjects were bothered by this interface design feature.

The goal of this initial study was not to comprehensively evaluate the displays, but rather to determine if the MCH-UVD was useful in determining display discrepancies and usable for operators (i.e., did people understand the process and level meanings.) The displays used in this case study, which were older versions of software currently in place, were redesigned prior to this experiment. Thus the MCH-UVD was able to identify (post-hoc) some discrepancies which were corrected after operational feedback. It should be noted that the fundamental goal of the MCH-UVD tool is to capture possible problems *before* deployment such that the quality improvement cycle is shortened, thus reducing training time and upgrade costs.

B. *Case Study 2*

The second study was a formal user experiment aimed at evaluating a new decision support tool designed to assist mission management and in-mission replanning during multiple UAV operations. The experimental task (reported in detail elsewhere [5]) involved subjects simultaneously supervising and updating the mission plans of four independent, highly autonomous UAVs with various levels of decision support assistance in a simulated task environment. Subjects' overall goal was to destroy a predetermined set of targets within a limited time period.

The experimental interface consisted of two displays: a tactical display and a decision support display (Fig. 6). The tactical display provided a vehicle interaction panel and a situation map which depicted the UAVs, scheduled routes, ground targets, and threat areas. The decision support display provided communication and status windows and a decision support window that provided UAV mission timelines, and in two of three study conditions, a 'StarVis' diagnostic and what-if tool (Fig. 6, zoomed view). The StarVis decision support tool displays existing scheduling problems and the potential problems that may arise if a certain rescheduling action is taken. In principle, the StarVis decision support tool should have supported in-mission replanning activities such as re-routing UAVs around emergent threat areas and revising UAV mission plans to include additional pop-up targets. The StarVis is a configural display, meant to specifically address the information integration problem noted in the discussion for Rating 8. See [5] for more details, but theoretically, users should have rated the display a 7 or lower since the display was designed to support direct perception-action.

The study involved 15 subjects performing several mission

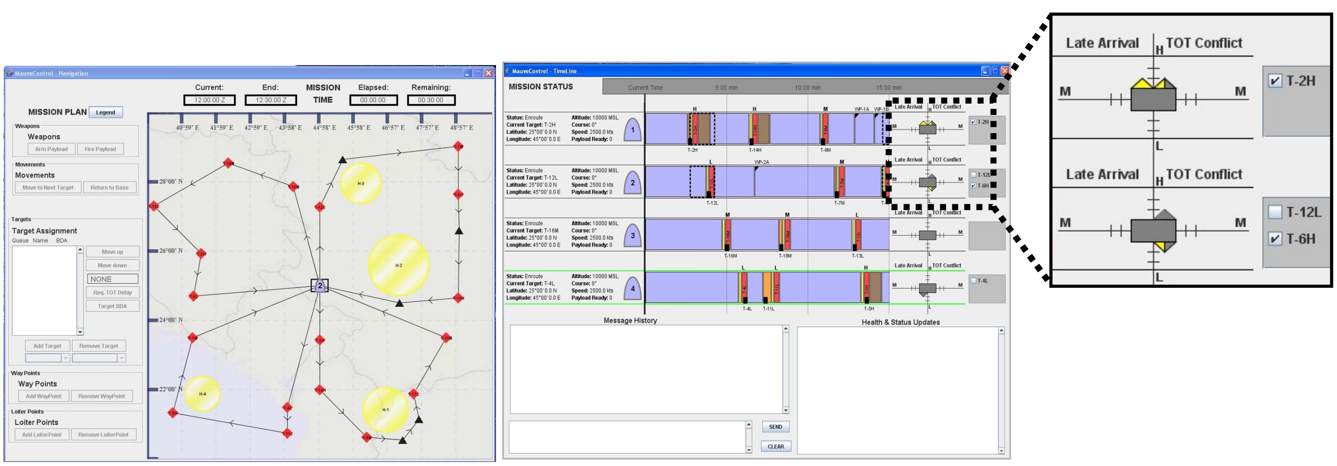

Fig. 6. The tactical (left) and decision support (right) displays of the multiple UAV supervision and replanning task, with a zoomed in view of the StarVis decision support tool.

trials. Subjects (also college students) were divided into three groups of five, each performing the mission trials with one level of decision support (timeline only, and two different StarVis variations). The performance results revealed that the best performers where those subjects that used the StarVis decision support, applied to individual UAVs as opposed to either those that had no StarVis or even those that had a global StarVis that applied to all UAVs. In terms of usability, these best performers generally thought the display met or exceeded the level 8 requirements originally intended in the design (MCH-UVD rating scale M=6.0, SD=3.0, MD=6; see Fig. 7). Four out of five of the best performers who completed the experimental task with the StarVis design submitted MCH-UVD ratings of 6 or below. While this rating indicates that users felt that information analysis was hampered by their inability to detect changes, they did feel that the StarVis design supported direct perception of integrated information. This supposition is confirmed by many of the comments subjects provided with their ratings, including: "because it does not display what would happen if one added or removed a target" (a change element).

Including the MCH-UVD evaluation tool in this user experiment helped us identify important deficiencies in the operator interfaces under investigation, even for the interface that produced the best mission performance. Despite the superior performance with the decision aid, there is clearly room for redesign, which should continue to improve performance. Without subjects' MCH-UVD ratings and accompanying rationale, we may have missed the fact that even an interface tool that produced relatively good mission performance needs improvements to truly support decision making in the time-critical UAV task environment.

Based on the interface deficiencies revealed by these MCH-UVD ratings, we are in currently in the process of redesigning the StarVis tool to better support information analysis and decision making for UAV replanning activities.

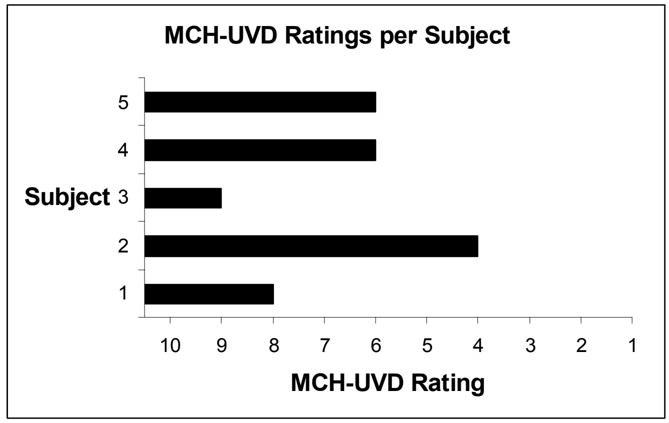

Fig. 7. Reported MCH-UVD ratings of the StarVis decision support tool on the multi-UAV operator interface.

VI. Conclusions

In summary, our limited experiences with the MCH-UVD evaluation tool so far show that it is a helpful tool for identifying deficiencies in the design of UV operator displays or display components. In particular, it helps to identify what level of information processing and decision support the interface provides to UV operators - activities critical to the success of most UV missions. However, as demonstrated by both studies, variation in opinion can be significant and subjective opinions should not necessarily guide interface design because often what users like can actually be detrimental to their performance [6]. Moreover, the numerical results from any application of the MCH-UVD should always be interpreted in light of users' comments.

These results are preliminary and the MCH-UVD scale now needs to be applied to UV displays in use by actual operators in order to determine the effectiveness of such a rating scale as well as to refine it. Other issues that need to be investigated further include examining how novice preferences (as represented in this study) differ from expert ratings and whether or not these ratings can be correlated to training time and performance.

The MCH-UVD tool was created as a more formalized and standard way for the designers of unmanned vehicle interfaces to evaluate their displays. Subjective ratings can be difficult to address because no design will ever please everyone. However, the MCH-UVD rating scale maps subjective ratings to specific design criteria and should help designers understand where in the information processing loop interface problems are occurring. By adopting this, or any other principled evaluation tool, companies and government agencies not only improve the overall user experience, but also can reduce the research and development cycle, as well as reduce the need for expensive operational upgrades.

Acknowledgments

This research was supported by AAI Corporation, CDL Systems Ltd., and MITRE Corporation.

References

[1] R. P. Harper, Jr., and G. E. Cooper, "Handling qualities and pilot evaluation " *Journal of Guidance, Control, and Dynamics*, vol. 9, pp. 515-529, 1986.

[2] P. F. Borchers, J. A. Franklin, and J. W. Fletcher, *Flight Research at Ames, 1940-1997: Fifty-Seven Years of Development and Validation of Aeronautical Technology (NASA SP-3300)* Moffett Field, CA: NASA, Ames Research Center, 1998.

[3] B. Shneiderman, *Designing the User-Interface: Strategies for Effective Human-Computer Interaction*, 3rd ed. Reading, MA: Addison Wesley Longman, 1998.

[4] CDL Systems, "Vehicle Control Station Style Guide VCS-540," Calgary 2005.

[5] A. Brzezinski and M. L. Cummings, "Decision Support Design for Workload Mitigation in Human Supervisory Control of Multiple Unmanned Aerial Vehicles HAL2006-05," MIT Humans and Automation Laboratory, Cambridge, MA 2006.

[6] A. Andre and C. Wickens, "When Users Want What's Not Best for Them," *Ergonomics in Design*, vol. October, 1995.