Enhancement and Application of a UAV Control Interface Evaluation Technique: Modified GEDIS-UAV

WENJUAN ZHANG, Dataminr Inc., New York, NY
DAVID FELTNER, United States Military Academy, West Point, NY
JAMES SHIRLEY, NAVAIR, New Bern, NC
DAVID KABER, University of Florida, Gainesville, FL
MANIDA S. NEUBERT, Khon Kaen University, Khon Kaen, Thailand

UAV supervisory control interfaces are important for safe operations and mission performance. We reviewed existing UAV interface design and evaluation tools and identified limitations. To address issues with existing methods, we developed an enhanced evaluation tool, the M-GEDIS-UAV. The tool includes detailed criteria for all aspects of UAV control interface design to support operator performance. It also supports quantitative and objective assessment of an interface. We prototyped three UAV information displays, including a digital control display, analog control display, and "massive" data display, as part of a simulated supervisory control interface. Six analysts, including three human factors experts and three novices evaluated the interfaces using the M-GEDIS-UAV. Inter-rater reliability was high for the human factors experts, suggesting training in usability analysis is necessary for tool application. Results also revealed the massive data display to produce significantly lower evaluation scores than the other displays. We concluded that the M-GEDIS-UAV was sensitive to interface manipulations and was most effectively used by human factors experts. Using the M-GEDIS-UAV tool can reveal the majority of design deviations from guidelines early in the design process toward increasing the effectiveness of control interfaces.

CCS Concepts: • **Human-centered computing** → **HCI design and evaluation methods**;

Additional Key Words and Phrases: UAV, interface design, usability assessment, human factors experts, GEDIS-UAV

ACM Reference format:
Wenjuan Zhang, David Feltner, James Shirley, David Kaber, and Manida S. Neubert. 2020. Enhancement and Application of a UAV Control Interface Evaluation Technique: Modified GEDIS-UAV. *ACM Trans. Hum.-Robot Interact.* 9, 2, Article 14 (January 2020), 20 pages.
https://doi.org/10.1145/3368943

This research was conducted when Wenjuan Zhang, David Felter, and James Shirley worked in the Human-Systems Engineering Lab at North Carolina State University. The research was supported by a grant from the National Aeronautics and Space Administration (NASA Grant No. NNX16AB23A). The views and opinions expressed are those of the authors and do not necessarily reflect the views of the NASA.

Authors' addresses: W. Zhang, Dataminr Inc., 6E 32nd St., 2nd Floor, New York, NY, 10016, USA; email: wzhang@dataminr.com; D. Feltner, Department of Behavioral Sciences and Leadership, United States Military Academy, West Point, NY, 10996, USA; email: david.feltner@westpoint.edu; J. Shirley, Department of Reliability and Maintainability Engineering, NAVAIR, New Bern, NC, 28560; D. Kaber, Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611, USA; email: dkaber@ufl.edu; M. S. Neubert, Program of Production Technology, Khon Kaen University, Khon Kaen, 40002, Thailand; email: manida@kku.ac.th.

2573-9522/2020/01-ART14
https://doi.org/10.1145/3368943

1 INTRODUCTION

Safe operation of Unmanned Aerial Vehicles (UAV) is the basis for successful mission completion and benefits of the technology. Unlike traditional manned-aircraft pilots operating in a three-dimensional world, UAV pilots rely on information presented through a ground control station (GCS). A well-designed interface is therefore critical to the effectiveness of task performance, such as vehicle control, navigation, environment hazard detection, and system health monitoring. As robotic technology is developed for more types of autonomous "behavior," human interfaces are used less for control and more for monitoring and diagnosis [10]. However, the importance of the operator interface should not diminish as the types of vehicle autonomy increase [10].

1.1 Human Factors Issues in UAV Accidents

Based on an analysis of UAV accident reports by Yesilbas and Cotter [36], approximately 60% of remotely piloted aircraft mishaps involved operation-related human casual factors. Another report by the Federal Aviation Administration (FAA) suggested human errors contributed to 21%–68% of UAV accidents [35]. For example, among 15 U.S. Air Force Predator accidents analyzed in this report [35], 8 of them were attributed to human factors issues. These issues included deficiencies of alerts and alarms design and various information presentation issues in head-up displays. Such human factors issues can be identified early in the systems design process to reduce or eliminate UAV damage or loss. To support accurate identification of human factors issues, a comprehensive evaluation tool for UAV remote control interfaces is needed.

1.2 Current Interface Evaluation Methods

A constrained review of UAV interface evaluation tools revealed several methods that have been applied in specific system evaluations to support human performance. A popular approach is usability testing, which has been used to assess whether interfaces present necessary functional features and if features are easy to use [7]. There are many different methods for testing usability but most measure learnability, efficiency, memorability, errors, and satisfaction to iteratively improve an interface [26]. This approach has been applied by a number of researchers in UAV interface evaluations as well as manned aircraft interface design analysis [17, 18]. For example, Irrizarry et al. [13] recruited participants to perform simulated tasks and rate usability of safety inspection drones. The feedback from participants suggested the need for better image quality and larger interface screens. As part of usability testing, surveys are often used to gather feedback from participants. The system usability scale (SUS) [17] is a survey commonly used in industry for evaluating general usability of an interface. The NASA-TLX (Task Load Index) [13] is a questionnaire often used in empirical studies to evaluate perceived cognitive workload from the perspectives of mental workload, physical workload, temporal workload, perceived effort, perceived performance, and frustration level. One drawback of usability testing is that user recruitment, data collection, and analysis for each interface iteration can be time-consuming and labor-intensive. As an alternative to testing, usability analysis can be performed via heuristic evaluations, where usability experts compare interface designs with established principles and identify discrepancies [22]. Due to this characteristic, usability analysis relies heavily on subjective opinion. Another limitation of heuristic evaluation is that this method does not directly recommend solutions to identified interface problems [25]. Beyond this, existing usability principles typically place more focus on traditional computer software dialogues, rather than complex supervisory monitoring systems.

Cognitive task analysis (CTA) is another established approach for interface evaluation. CTA is an extension of task analysis techniques and is used to develop an understanding of human knowledge, thought processes, and goal structures in task performance [14, 33]. CTA modeling has been applied to tasks with interactive systems. For example, Kaber et al. [15] explored using goal-directed task analysis (GDTA) along with abstraction hierarchy (AH) modeling to characterize the user knowledge requirements of a supervisory interface for high-throughput biological screening processes. The results proved useful for improving interface intuitiveness and interactivity through redesign processes [19]. However, considerable application time and high complexity of CTA models are major challenges for this approach [15, 19]. Furthermore, CTA methods restrict interface evaluation to pre-defined tasks. Model adjustments are necessary when there are changes in the operating procedure (e.g., additional tasks, altered steps). More efficient approaches are necessary to support interface design in the UAV industry.

In addition to these traditional assessment techniques, some researchers have attempted to adapt established tools from other domains to the UAV domain. The FAA has used the Cooper-Harper subjective workload rating scale [12] for many years to assess manned aircraft pilot cognitive demands and to gauge whether a cockpit interface design is effective from a performance perspective. Cummings et al. [4] modified the Cooper-Harper scale to the unmanned vehicle (UV) domain with a focus on evaluating how well control displays support basic operator information processing. The new tool (Modified Cooper-Harper for Unmanned Vehicle Displays; MCH-UVD) followed the structure of the MCH by implementing a 10-point uni-dimensional rating scale on information processing support. Donmez et al. [7] conducted experiments involving application of the MCH-UVD. Pilots executed simulated UAV search and rescue tasks and were asked to evaluate vehicle control interfaces with the MCH-UVD subsequent to experiment trials. While results suggested the method to be predictive of operator performance, it should be noted here that superior task performance is not necessarily a comprehensive and reliable indicator of usable interface design. Operator performance can also be influenced by experience level, task complexity, and luck. In both the Cummings et al. and Donmez et al. [4, 7] studies, there was no comparison of performance or evaluation results made across different interfaces. Moreover, few (if any) other studies have applied the MCH-UVD for unmanned systems interface evaluation. Therefore, the validity of MCH-UVD as an interface evaluation tool still needs to be further investigated with comparison across different interfaces.

1.3 GEDIS and GEDIS-UAV

Ponsa and Díaz [31] developed a cognitive ergonomic guideline for supervisory control interface design with the objective of improving the efficiency of human-machine systems as part of semi-automated industrial processes. They reviewed human interface design guidelines related to automation control and supervision tasks (see Table 1) and proposed the "ergonomic guideline for supervisory control interface design" (GEDIS; Guia ergonómica para el disenõ de interfaz de supervision (in Spanish)). The GEDIS consists of 10 key design indicators that cover different aspects of interface design for process supervisory control, including: the interface architecture, information distribution, interface navigation, display colors, text fonts, device status displays, process value displays, graphs and tables, data-entry commands, and alarms. Each design indicator has a substructure, including identification of a diverse set of design characteristics or "sub-indicators." Any supervisory control interface can be quantitatively evaluated based on the degree of design satisfaction of the sub-indicators (i.e., characteristics of interface features). A rating scale ranging from 0 = "not appropriate" to 5 = "appropriate" with an intermediate value of "acceptable/moderate" is used for this purpose. As an example, the design indicator of interface "navigation" can be evaluated based on two sub-indicators, including: (1) the relationship navigability with interface architecture, and (2) navigability between screens. Each of the two sub-indicators can be assigned a score of 5, 3 or 0, depending upon whether each design characteristic is considered "appropriate," "acceptable," or "not appropriate" for supporting human use. In some cases, a score is assigned

Table 1. Guidelines Used as Basis for Interface Evaluation Tool Development

Guidelines Reviewed by GEDIS	Guidelines Reviewed by GEDIS-UAV
ISO 11064-7:2006 (Ergonomic design of control centers—Part 7: Principles for the evaluation of control centers) [40]	ARINC 601—Aims to normalize the definition of a Cockpit Display System (CDS)
Human Factors Design Standard (HFDS)	STANAG 4586 [3]
NUREG 0700: Human Interface Design Review Guidelines [29]	DO-178B—Software Considerations in Airborne Systems and Equipment Certification [38]
I-002 Safety and Automation Systems NORSOK	Joint Architecture for Unmanned Systems (JAUS)—JAUS Unmanned Ground Vehicle Service Set
Man Systems Integration Standard (NASA-STD-3000)	Joint Architecture for Unmanned Systems (JAUS)—JAUS HMI Service Set
	ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs)—Part 11: Guidance on usability

simply based on the presence of certain features. For instance, with respect to the interface "architecture," a sub-indicator of "existence of maps" can only be assigned a score of 5 or 0, indicating that an interface map is present or not. The score of a design indicator is simply the weighted average of all sub-indicator scores (Equation (1)):

$$Indicator\ Score = \frac{\sum_j^J w_j Subindicator_j}{\sum_j^J w_j}, \tag{1}$$

where $\mathcal{J}$ = number of sub-indicators, and w_j = weight of the jth sub-indicator.

Similarly, a global evaluation index for the supervisory control interface can be obtained by calculating the weighted average of indicator scores (Equation (2)). Weighting factors can be determined based on expert opinion of the relevance of any given design indicator/feature to human information processing in the target interface application:

$$Global\ Evaluation\ Index\ \ (GEI) = \frac{\sum_i^I p_i Indicator_i}{\sum_i^I p_i}, \tag{2}$$

where I = number of indicators, and p_i = weight of the ith indicator.

Ponsa Asensio et al. [30] and Ponsa and Díaz [31] applied the GEDIS to a Sugar Technology Center simulator interface and generated design recommendations to achieve a maximum GEDIS Global Evaluation Index (GEI), which was considered to represent optimal interface design. It was suggested that the GEI of an interface should have an initial value of 3–4 and designers should ultimately work to achieve a score of 5 with the help of the GEDIS checklist [31].

Lorite et al. [21] adapted the original GEDIS for application to evaluation of UAV graphical user interfaces (GUIs). In their effort to create a GEDIS-UAV evaluation tool, the researchers reviewed standards and guidelines related to the design of GUIs for UAVs (also see Table 1). The desired design features/indicators were carried forward from the original GEDIS, however, modifications were made to desired sub-indicators (feature characteristics) to accommodate UAV interface technology. For example, under the indicator of interface "architecture," sub-indicators, including "number of display levels" and "existence of map" were removed, and "number of screens" was added. The GEDIS-UAV inherited the scoring system of the GEDIS, where the GEI and indicator scores remain weighted averages.

The GEDIS-UAV was subsequently used to evaluate a UAV ground control interface, which was designed based on the ARINC 661 standard, comprising a display mounted on a control panel and similar to a conventional aircraft cockpit display [21]. The initial GEI for the interface was 3.8. A set of design changes were proposed based on the evaluation results and design indicator scores, including redistributing components of the interface, improving data visualization, improving alarm content and adjusting display font sizes. These modifications brought the GEI to 5, which is the maximum value of the rating scale. The researchers expected the enhanced interface to minimize the possibility of human error in use [21]. However, they did not conduct a follow-up study to test the benefits of the enhanced design, based on application of the GEDIS-UAV tool.

1.4 Issues and Need for Enhancement

Although various usability and workload evaluation methods have been applied and adapted for UAV interface evaluation, most existing tools are not designed to assess interface capability to support UAV functions [4]. The GEDIS-UAV provides a good starting point for including both functionality and usability considerations in UAV interface evaluations. It also provides quantitative assessment of degree of conformance of specific interface designs with existing guidelines and allows quick comparisons among different designs. However, there are a few issues related to application of the GEDIS-UAV.

First, the original tool did not provide a basis or justification for selection of desired interface design indicators/features. This drawback resulted in confounded indicators and an absence of some important UAV interface indicators, such as maps and navigability features. An overarching usability and functionality framework is needed for indicator identification and selection, such as Nielsen and Molich's [27] usability principles, to ensure no missing or confounded indicators. The second limitation of the GEDIS-UAV is that the selection of sub-indicators (i.e., characteristics of design features) was not supported by detailed references to design standards or guidelines. The current set of sub-indicators under each indicator was not a complete representation of all characteristics related to contemporary UAV interface design components/features. This issue also resulted in overlap of sub-indicators. For example, the "graphs and tables" design indicator includes sub-indicators of "format," "visibility," "location," and "grouping" as bases for interface evaluation. However, these are not independent design features and can be influenced by each other. A third limitation of the GEDIS-UAV tool is that the method of assessment (rating) of each interface sub-indicator (i.e., appropriate, moderate, not appropriate) is not objective. For instance, in the assessment of appropriate font combinations, the answer could vary depending on the analyst's eyesight, age, cultural background, reading preference or sensitivity to font characteristics. As another example, assigning a score to the sub-indicator "format" under "graphs and tables" is entirely subjective. Interface evaluation following this approach can be highly vulnerable to personal preferences and rater emotional states [37], instead of being an objective indicator of how well a design supports usability and performance. As a fourth limitation, the language used in presenting many of the sub-indicators as part of the GEDIS-UAV tool introduce ambiguity and could lead analysts to make incorrect judgments about specific design feature characteristics. For example, the GEDIS-UAV inherited the indicator of interface "navigation" from the GEDIS tool, which refers to "the logical linkage of interface items for supervisory control of industrial automation." However, when it comes to the UAV domain, interface navigation has a completely different meaning; that is, planning and controlling a UAV route vs. monitoring process flow channels and pump settings. Another example of confusing language among GEDIS-UAV sub-indicators can be found in the design guideline for, "relationship with text." It is unclear from the word "relationship" that this guideline refers to how color influences the presentation of text and what display colors are appropriate for interface design.

Table 2. Indicators in M-GEDIS-UAV

Usability Related	**Functionality Related**
Display Layout (DL)	Map and Navigation (MN)
Information Presentation (IP)	Status and Devices (SD)
Color (C)	Data Entry Command (DEC)
Text (T)	Alarm (A)
Physical Control (PC)	

The lack of comprehensiveness and objectivity in current assessment approaches makes it difficult to make comparisons across multiple candidate UAV interface designs as part of a systems design process. In an attempt to address these issues, the present study developed a Modified GEDIS-UAV (M-GEDIS-UAV) and conducted follow-up experiments to assess the validity and sensitivity of this new tool.

2 DEVELOPMENT OF THE MODIFIED GEDIS-UAV (M-GEDIS-UAV)

2.1 Enhancement on Indicators and Sub-indicators

To address limitations of the GEDIS-UAV, we took both a "top-down" and "bottom-up" approach to identifying and organizing interface design indicators as well as specification of design characteristics for interface evaluation. The top-down approach involved identifying a collection of usability and functionality principles and features as bases for organizing desired design features to be addressed by the new M-GEDIS-UAV. The usability principles were based on usability heuristics [27, 28, 34] and the functionality features were based on functional interface components observed from various UAV studies [1, 2, 4, 8, 10, 21, 22, 24]. The structure of the original GEDIS-UAV was also used as a reference for organizing the indicators. Following the top-down approach, a modified set of desired interface design indicators were identified for the M-GEDIS-UAV, as shown in Table 2. With this set of indicators, the M-GEDIS-UAV is appropriate for evaluating UAV supervisory control interfaces with a computer interface. The tool does not take into consideration exploratory UAV control technologies such as virtual reality (VR) and gesture-based controls.

The bottom-up approach involved compiling an elaborate list of design sub-indicators (feature characteristics) from existing standards. All industrial guidelines listed in Table 1 were reviewed as part of this approach. We also reviewed additional guidelines related to supervisory control interface, including the Unmanned Aircraft System Ground Control System Human Machine Interaction (UAS GCS HMI) reference [20] and Military-Standard-1472 [39]. A subset of guidelines was selected for characterizing each interface design indicator (resulting from the top-down approach). This process was facilitated by using a set of selection rules, including: (1) the guideline is applicable to the UAV domain; (2) the guideline addresses design issues relevant to UAV supervisory control interfaces (and not robotic systems, in general); (3) the guideline can be applied to contemporary technologies used in UAV supervisory control interfaces (e.g., guidelines for cathode ray tube displays were not included); and (4) the guideline is sufficiently descriptive to allow for assessment of the degree to which a particular interface design is conforming or nonconforming with the guideline. If multiple guidelines identified through the bottom-up approach described the same interface content, then they were combined as one criterion. If there were conflicts among guidelines, then the guideline referring to an interface use scenario closest to UAV operation was adopted. Among all the guidelines selected for assessment of each interface design indicator/feature/component, those addressing the same (or similar) interface feature characteristics were grouped under a sub-indicator. The sub-indicators were named based on the design

Table 3. Indicators in Modified GEDIS-UAV

Indicator	Subindicators	Indicator	Subindicators
Display Layout (DL)	Screen Location	Status and Devices (SD)	Indication of System State
	Readability		Analog / Digital Displays
	Information Density		Scales
	Controls		Numbers Displayed
	Menu Structure	Data Entry (DE)	Cursor
	Windows		Fields
Information Presentation (IP)	Hierarchy & Grouping of Info		Intra-field Separators
	Labels & Titles		Field Labels
	Use of Coding		Units
	Shape Coding		Data Entry Validation
	Color Coding		User Aids
	Alphanumeric Coding		Data Selection
	Line Coding	Alarm (A)	Alarm Message Style
Color (C)	Color Discrimination		Alarm Content
	Color Luminance		Alarm Handling
	Color Contrast		Auditory Signal
	Color Use		Alarm Priorities & Grouping
	Brightness		Alarm Color
Text (T)	Font		Alarm State
	Spacing		Flash
	Abbreviations	Physical Control (PC)	Keyboards
	Character Luminance		Fixed-function Keys
	Character Contrast		Pointing Devices
	Capitalization		Mice
	Text Use		Joystick and Trackball
	Underlining		Alternate Input Devices
	Bold		Interchangeability
	Table		Hand-Operated Displacement Joystick
Map & Navigation (MN)	Visibility		Finger-Operated Displacement Joystick
	Overlays and Map Items		
	Map Colors		
	Labels and Symbols		
	User Preferences		

characteristics commonly referenced in the guidelines. The guidelines under each sub-indicator serve as criteria for evaluation of interface conformance with each sub-indicator (i.e., design characteristic). All subindicators of the M-GEDIS-UAV are summarized in Table 3.

After all relevant criteria were identified from the existing UAV and human factors design standards, the phrasing of each design characteristic and the meanings were edited for clarity. These steps were intended to ensure no conflicts among criteria and that all criteria referred to one (and only one) design characteristic. Another purpose of the editing was to support analyst ease of interpretation and determination of whether a specific design was conforming or nonconforming. Currently the M-GEDIS-UAV contains 290 unique criteria for comprehensive evaluation of UAV supervisory interface design characteristics. The guidelines in M-GEDIS-UAV refer

Table 4. Evaluation Criteria for “Menu Structure”

Menu Structure (MS)	The interface provides an appropriate maximum number of options for different types of graphical controls: (a) Radio buttons: 1–6 options; (b) Static Menus: 3–10 options; (c) Menu Bars: < 10 options; and (d) Scrolling Menus: >10 options.
	The number of selections required to reach the desired option in complex menus is no more than four steps.
	When a user selects a menu option and no computer response is immediately observable, the software provides some other acknowledgment of the selection.
	Menu options are presented in a single vertical column, aligned and left justified (exception: menu bars).
	Destructive commands (e.g., delete, exit) are placed at the bottom of menus.
	Options for opposing actions (e.g., save and delete) are not placed adjacent to each other.
	Primary windows’ menu bars extend the full width of the primary window.
	System menus include the following options: end a session, review system status, define user preferences, manage alerts, and change a password.

to observable design features and characteristics that can be compared with criteria statements, allowing for objective assessment of an interface. As an example, Table 4 presents the criteria for “Menu Structure,” which is a subindicator under the indicator “Display Layout.” For each design indicator addressed by the M-GEDIS-UAV, we prepared a separate spreadsheet presenting its sub-indicators and their design criteria. The sources for all the guidelines are also identified in the spreadsheets. The complete evaluation spreadsheets are available at https://www.ise.ufl.edu/kaber/publications/supplemental-information-for-publications/.

2.2 Enhancement on Scoring Mechanism

The M-GEDIS-UAV has a checklist structure. A set of design criteria is presented in the form of a checklist for each sub-indicator and provides a basis for objective assessment of interface designs. To evaluate a design sub-indicator with the M-GEDIS-UAV tool, an analyst compares the actual interface design characteristics with related criteria and determines whether each criterion is satisfied or not. That is, the analyst assigns a binary score (“1” if satisfied, “0” if not) for each criterion. If the characteristic described by a criterion is not related to the interface under evaluation, then the analyst assigns a response of not applicable (“NA”) and the criteria is not considered in scoring of the interface design. In this way, the M-GEDIS-UAV eliminates the need for an analyst to provide a subjective rating of each interface design characteristic with multiple arbitrary scores (i.e., 0, 3, 5). The approach also eliminates the need for justification or rationale for different ratings on characteristics. Instead, objectivity of the assessment is maximized by matter-of-fact comparisons against individual design guidelines, each of which are supported by extensive design literature. The score for each sub-indicator can be determined by averaging the scores for the criteria within the sub-indicator checklist.

On this basis, an indicator score becomes the average of all relevant sub-indicator scores. That is, for each desired design feature, the degree of interface conformance (a percentage) with guidelines for specific characteristics of the feature/component is calculated. Subsequently, the GEI is calculated as the average percent satisfaction of desired design features for an interface under evaluation. Figure 1 shows an example of evaluation for the "Status and Devices" indicator. The GEI is simply obtained by averaging all the indicator "scores." The GEI can vary between 0 and 1, with 1 representing an optimal design. The GEI score represents the extent to which an interface is aligned with recognized design guidelines and can be used as a reliable usability indicator for UAV interfaces. At this stage of development of the M-GEDIS-UAV, we assumed that all design indicators contribute equally to the usability and functionality of the interface design. However, the M-GEDIS-UAV scoring system allows for assignment of weighting factors to design indicators, if designer or user expertise is available to support identification of weighting factors.

3 INITIAL APPLICATION OF M-GEDIS-UAV

As a preliminary assessment of the usefulness of the M-GEDIS-UAV tool, we applied the tool to the Multi-attribute (flight) Task Battery II (MATB-II [41]; see Figure 2 for an image of the MATB-II interface). The purpose of this evaluation was to determine if the structure and language of the M-GEDIS-UAV was clear for analysts and if the tool could be used to identify deviations of aviation display design from identified criteria. The MATB-II interface simulates a number of flight tasks that a pilot may perform (e.g., aircraft system status monitoring, target tracking, radio communication management, etc.).

3.1 Experiment

Three of the authors worked as analysts for the preliminary applications of the M-GEDIS-UAV. All three analysts completed at least two years of human-factors-related coursework, which allows them to fully understand the evaluation criteria as part of the tool. Therefore, they were considered as human-factors expert analysts. None of the analysts had previous experience using the MATB-II interface. Prior to the interface evaluation, an experienced user of MATB-II trained the analysts on the interface components and functionalities. The analysts also reviewed videos of the interface being used to perform each flight task. Each analyst then conducted an independent evaluation of the MATB-II interface with the M-GEDIS-UAV. They were also given access to a functional version of the interface for thorough assessment of the degree of design conformance with the M-GEDIS-UAV criteria. The analysts were asked to provide feedback on the format, comprehensiveness, and clarity of criteria language as part of the tool. Results of this preliminary application were used for additional improvement of clarity of language in the M-GEDIS-UAV.

3.2 Results and Discussion

The M-GEDIS-UAV revealed relatively low but consistent analyst scores for the MATB-II interface ranging from 74–78% conformance with all applicable design criteria. Intra-class correlation coefficient (ICC) were used to assess the degree of inter-rater reliability in applying the tool. The ICCs were calculated using the two-way agreement single model, as described by McGraw and Wong [23]. Results revealed a moderate (ICC = 0.51 [32]) and significant ($F(7, 14.1) = 3.74$, $p = 0.017$) correlation among the analysts' ratings of the various (9) interface design indicators. However, we found that there was a high degree of variance among analyst scores for the indicators of display "color," "status and device" displays, and "alarm" design. The analysts' feedback also indicated some concern with the clarity of phrasing of criteria on these interface design features as well as some criteria referring to multiple design characteristics. Based on these results, we revisited the design criteria checklists to ensure the language was clear and concise. Across all design features (e.g., color, etc.), descriptions of criteria were simplified and clarified for easier

Evaluate Indicator: Status and Devices (SD)

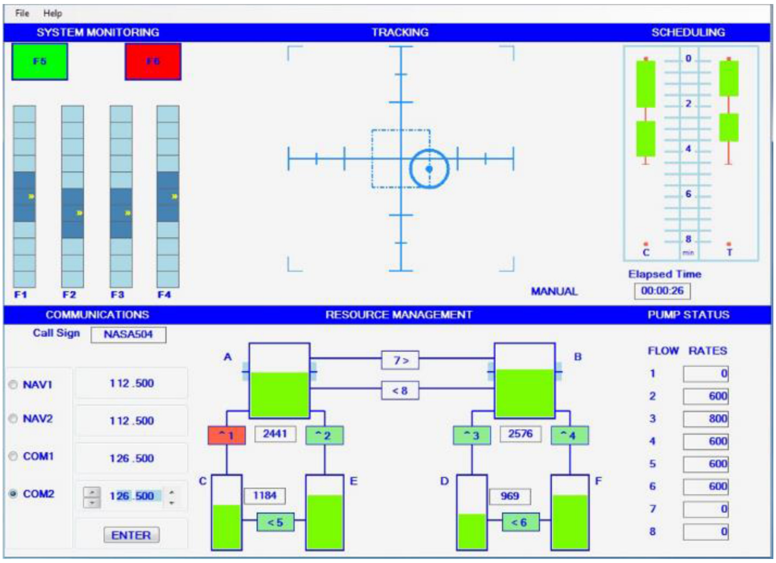
Indicator score: 83%

Enter "1" under the "Conformance" column if the interface conforms to the specific design criterion. Enter "NA" if the criteria is not applicable.

If all inputs for a subindicator are "NA", enter "NA" for the "Subindicator score".

Subindicator score	Subindicators	Conformance	ID	Criteria related to this sub-indicator	Reference	Comments
80%	**Indication of System State (ISS)**	1	**ISS1**	Display provides a positive and unambiguous indications of system state (e.g., indicating "power on" as indicated by a blinking cursor).	NASA-STD-3000: 9.4.2.3.2	
		0	**ISS2**	Positive indications (e.g., on, active, run, etc.) are used consistently throughout the interface.	NASA-STD-3000: 9.4.2.3.2	
		1	**ISS3**	Status icons and symbols are easily discriminable from all other icons and symbols.	NUREG 0700: 1.3.4	
		1	**ISS4**	Status icons are assigned unique meanings and used consistently throughout all applications.	NUREG 0700: 1.3.4	
		1	**ISS5**	Status symbols are assigned unique meanings and used consistently throughout all applications.	HFDS 8.3.3.1.2	
50%	**Analog / Digital Displays (ADD)**	1	**ADD1**	Digital displays are used for precise readings of quantitative values, where trend information is not needed.	HFDS 6.2.1.2	
		0	**ADD2**	Analog displays are used to present values in relation to ranges or zones, when users require trend information.	HFDS 6.2.1.2	
100%	**Scales (Sc)**	NA	**Sc1**	A moving pointer scale is used when: 1: Numbers and the scale need not be read and pointer position change is easily detected. 2: There is a simple and direct relation of motion of the pointer to the motion of a setting knob. 3. Any pointer position change aids in system monitoring. 4: Tracking has the simplest relation to manual control motion.	HFDS 6.2.6.1	
		1	**Sc2**	Scale graduations progress by 1, 2, or multiples of 5 units or decimal multiples thereof.	HFDS 6.2.6.1.4	
		1	**Sc3**	Scales have fewer than 12 major scale divisions.	HFDS 8.3.3.1.16	
		1	**Sc4**	Scale major divisions have fewer than 10 sub-divisions.	HFDS 8.3.3.1.16	
		NA	**Sc5**	For one-revolution circular scales, zero should be at 7 o'clock, and the maximum value should be at 5 o'clock, with a 60-degree break in the arc.	HFDS 8.3.3.1.12	
86%	**Numbers Displayed (ND)**	1	**ND1**	Numeric data is displayed in decimal form rather than binary, octal, hexadecimal, or other number system.	HFDS 8.2.7.1	
		1	**ND2**	There are no leading zeros in numeric entries for whole numbers.	HFDS 8.2.7.2	
		0	**ND3**	Integers are right justified.	HFDS 8.2.7.3	
		1	**ND4**	The system does not require the entry of the decimal point at the end of an integer.	HFDS 8.2.7.4	
		1	**ND5**	Displayed values contain the number of significant digits required for users to perform their tasks.	HFDS 8.2.7.5	
		1	**ND6**	Numeric displays accommodate the full range of the variable.	HFDS 8.2.7.6	
		1	**ND7**	All numbers are oriented upright.	HFDS 8.2.7.7	
NA	**Graphs (G)** **[If there is no graph present, please consider this subindicator NA]**	NA	**G1**	Graphs convey enough information to interpret without referring to additional sources.	NUREG 0700: 1.2.5	
		NA	**G2**	Graph data and graph elements (color, labels, legends, etc.) are clearly labeled.	NUREG 0700: 1.2.5	
		NA	**G3**	Graph data and elements are unobstructive to data readibility.	HFDS 8.3.3	
		NA	**G4**	Graphs have only one (1) scale on each axis.	HFDS 8.3.3.1.6	
		NA	**G5**	When cyclic data is displayed, at least one full cycle is presented.	HFDS 8.3.3.3.7	
		NA	**G6**	Curves representing values projected beyond the actual data set are coded distinctly from curves representing actual data.	HFDS 8.3.3.3.6	
		NA	**G7**	When a graph's emphasis is on the area between two (2) curves, that area is filled with color or a pattern.	HFDS 8.3.3.4.1	
NA	**Charts (Ch)** **[If there is no chart present, please consider this subindicator NA]**	NA	**Ch1**	Pie charts are used to show the proportional distribution of categories with respect to the sum of the categories.	HFDS 8.3.3.7	
		NA	**Ch2**	Flowcharts are used for showing schematic representations of sequential processes.	HFDS 8.3.3.11	
100%	**Controls (CTL)**	1	**CTL1**	Each type of control in an application is consistent and visually distinct.	HFDS 8.13	
		1	**CTL2**	Pointers are distinguishable under all light conditions on all parts of the display.	HFDS 6.2.6.1.16	

Fig. 1. Subindicators and evaluation criteria for the "Status and Devices" indicator, with example ratings and aggregated scores.



Elapsed Time 00:00:26

Call Sign NASA504

○ NAV1 112.500
○ NAV2 112.500
○ COM1 126.500
● COM2 126.500

2441
2576
1184
969

FLOW	RATES
1	0
2	600
3	800
4	600
5	600
6	600
7	0
8	0

Fig. 2. The MATB-II interface (top left—system monitoring task; top center—tracking task; top right—scheduling task; bottom left—communication task; bottom center and right—resource management task).

analyst interpretation. These modifications also included removing any "fully loaded criteria" (i.e., guidelines that make reference to more than one aspect of display design).

4 APPLICATION OF M-GEDIS-UAV TO UAV MISSION PLANNER INTERFACE

Subsequent to further modifications of the M-GEDIS-UAV, the tool was applied to the ArduPilot (UAV) Mission Planner (MP) interface to assess usefulness for identifying interface design deficiencies as well as sensitivity to interface feature manipulations. The MP is a full-featured GCS application that provides a graphical user interface for UAV supervisory control. This interface was used because it contains common components of UAV interfaces observed in various UAV studies [1, 2, 4, 8, 10, 21, 22, 24] and the software is freely available online. The interface consists of four sections (see Figure 3), with each providing a unique set of features. The upper-left section presents a simulated primary flight display (PFD) with horizon line. This display contains typical system information that would be commonly found in manned aircraft. The upper-right section is a geographical projection of the simulated UAV's environment (via Google Earth), an image representing the UAV itself, its projected course, and a visual indication of upcoming flightpath waypoints. The lower-left section is a configurable auxiliary box. A set of selectable tabs allows this box to toggle between a system information display and immediate action command interface for the UAV. The system information can be displayed with a few options, including digital or analog displays (see Figure 4). Finally, the lower-right portion of the interface is an active mission planning tool that allows an operator to enter future commands into a UAV's flight plan. Unlike the immediate actions available in the auxiliary box, this section is focused on long-term operation of the vehicle and serves as a proactive tool for mission changes.

Each section of the MP interface corresponds to one or multiple feature sets that can be evaluated with the M-GEDIS-UAV tool. For example, the system information display would be most appropriately evaluated using the M-GEDIS-UAV criteria for Status and Devices indicators. Some guidelines under other indicators might be affected by the analog gauges but the display (itself) would have the most significant impact on the score for Status and Devices indicators.

WP Radius 1.9999
Default Alt 50
Relative
☐ Verify Height
Alt Warn 10
☐ Spline

	Command	Dela				Lat	Long	Alt	Delete	Up	Down	Grad %	Angle	Dist	AZ
1	TAKEOFF	0	0	0	0	0	0	49.96719	X			0	0	0	0
2	WAYPOINT	0	0	0	0	-35.362124	149.165066	49.96719	X			11.9	6.8	418.3	353
3	WAYPOINT	0	0	0	0	-35.356935	149.163925	49.96719	X			0.0	0.0	1923.2	350
4	LOITER_TIME	60	0	1	0	-35.3569962	149.163866	29.98688	X			-70.4	-35.1	28.4	218
5	WAYPOINT	0	0	0	0	-35.35644	149.165019	49.96719	X			5.0	2.9	398.6	59
6	LOITER_UNLIM	0	0	1	0	-35.3564012	149.1649818	49.96719	X			0.0	0.0	18.0	322
7	WAYPOINT	0	0	0	0	-35.356129	149.166601	49.96719	X			0.0	0.0	491.9	78
8	WAYPOINT	0	0	0	0	-35.363226	149.167843	75	X			1.0	0.5	2615.3	172
9	LOITER_UNLIM	0	0	1	0	-35.363226	149.1677283	49.96719	X			-73.4	-36.3	34.1	270

Lat -35.3632621
Long 149.1652373
Alt (abs) 6287.093223
GEO -35.363263 149.165238 0.00m
☐ Tuning ☐ Auto Pan Zoom 15.0

Fig. 3. MP interface: Primary Flight Display (upper left), auxiliary box (lower left), system projection (upper right), flight planner (lower right).

AirSpeed (knots) 0.00
Dist Traveled (ft) 0.00
GroundCourse (deg) 0.00
Altitude (ft) 0.00
Bat Remaining (%) 100.00
linkqualitygcs 100.00

Fig. 4. System information display (digital display—left; analog display—right) within the auxiliary box of MP interface.

4.1 Experiment 1: Preliminary Application

4.1.1 Interface Analysts. Two groups of interface analysts evaluated the MP interface with two different system information display types, including digital vs. analog. The first group of analysts were the same from the preliminary M-GEDIS-UAV application. They were familiar with the M-GEDIS-UAV and had sufficient human factors background to fully understand the criteria. This group (referred as "human factors expert" group) had opportunity to explore the interface and functionalities of the MP prior to the experiment. The other group was a convenience sample of three novice analysts recruited from a master's program of the local university. The novice analysts had very limited coursework (i.e., less than 6 months) in the human factors area. They had no previous knowledge of the M-GEDIS-UAV tool or the MP interface. Some training on the M-GEDIS-UAV tool was provided for the novice group. After receiving explanation on the structure of M-GEDIS-UAV, the novice analysts practiced application of the tool by evaluating the MATB-II

Table 5. GEIs Across Displays and Analyst Expertise Levels: Mean (SD)

	Analog Display	Simple Digital Display
Novice	0.81 (0.10)	0.83 (0.09)
Expert	0.76 (0.06)	0.77 (0.08)

interface for a single design indicator (Color). The training for novice analysts lasted approximately 45 min. The training was intended to familiarize the novices with the process of applying the tool. A full evaluation of the MATB-II was not permitted so that the novices had limited prior knowledge of the specific interface design guidelines or functions of a UAV control interface. The objective here was to assess the potential importance of human factors expertise in application of the tool. Both groups were required to manually record their evaluation of the MP interface on a print-out of the M-GEDIS-UAV indicator spreadsheets, which were subsequently transcribed into digital format by a researcher.

4.1.2 Evaluation Procedure. This evaluation involved the digital and analog versions of the MP interface. To ensure consistent exposure to interface dynamics, each analyst was shown two videos of an operator using the MP interface to perform a set of generic UAV control operations. The first video was a simulated semi-autonomous mission with the UAV following a simple flightpath. At specified times and vehicle locations, the operator was instructed to enter new waypoints or commands into the GCS or to modify existing waypoint specifications. Throughout the mission, the operator was expected to monitor the UAV's speed, altitude, and other secondary parameters. The second video also presented interface alarms caused by operator errors in control and showed actions of acknowledging alarms. The analysts were provided with a description of the UAV control tasks presented in the videos but there were no narrative or verbal protocols as part of the recordings. The operations in the videos were based on responsibilities of a UAV pilot, specifically to: (1) control and monitor the location and flight path of the aircraft; and (2) recognize and respond to off-nominal conditions. The scenarios were reviewed by an Army officer with extensive experience in UAV operations to ensure that they were representative of real-world situations. The interface evaluation took approximate 2 hours for expert analysts and 2.5 hours for novice analysts.

Since the system information display only represents a small portion of the interface, we expected the display manipulation (i.e., digital or analog format) to only impact a few interface design indicators. Based on their familiarity with the M-GEDIS-UAV and the interface features, the expert analysts identified the impacted indicators to include: "Color," "Text," and "Status and Devices." The analysts were required to perform a complete evaluation of the simple digital display followed by an evaluation of the analog display in terms of the three impacted indicators.

4.1.3 Results and Discussion. The GEIs for the analog and digital displays for the two groups of analysts are summarized in Table 5. Diagnostics on the GEI scores revealed violations of normality as a basis for parametric analysis. Therefore, rank transformation was applied to the scores, which were then submitted to an analysis of variance (ANOVA) to yield a non-parametric test. Expertise level and interface type were considered as independent variables in the model. Analyst was included in the model to account for inter-rater variability, which was nested within expertise level. Results revealed a statistical significant effect of "expertise level" ($F(1,5) = 8.13$, $p = 0.034$) and "analyst (expertise level)" ($F(4,5) = 16.61$, $p = 0.004$). Although the digital display produced slightly higher GEIs than the analog display, the effect of display type was not statistically significant ($F(1,5) = 2.36$, $p = 0.185$). The lack of sensitivity is likely due to variability in scores among analysts. Another reason possible reason is that the difference between the digital and analog display was limited to numeric vs. round-dial and pointer gauges. This difference only appeared in one section of the interface (i.e., the auxiliary box) and was not sufficient to impact the GEI for the entire MP interface.

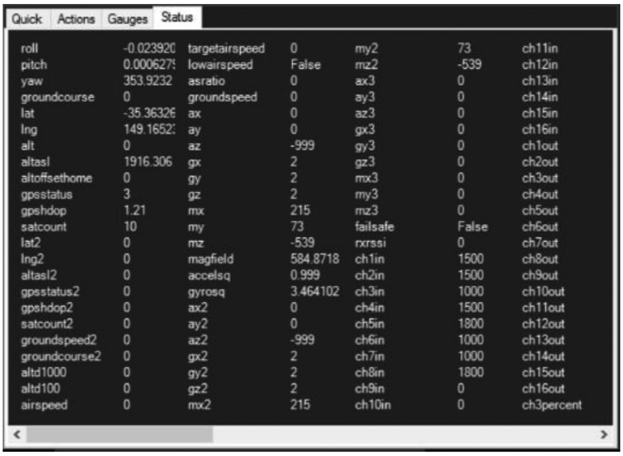
As in the preliminary assessment of the M-GEDIS-UAV, inter-rater reliability in use of the tool was analyzed with ICCs calculated on the design indicator scores for both the novice and expert groups. The novice group's ICC was 0.204, representing poor reliability according to Portney and Watkins [32]. The expert group showed moderate reliability (ICC = 0.429 [32]) and a significant correlation ($F(8, 15) = 4.01$, $p = 0.01$) among evaluation results. As an additional point of reference, an ICC $>= 0.7$ is considered to be acceptable as a basis for clinical practice [5]. It is possible that novice analysts were not able to identify as many design guideline violations as experts due to limited knowledge of the tool and interface. It was also noted that novice analysts produced more "NA" responses than judgments of "sufficient" or "not met" across all subindicators. On average, the novice analysts rated "NA" on 150 criteria out of 290 criteria as part of the M-GEDIS-UAV, as compared to expert "NA" rating for, on average, 111 out of 290 criteria.

Another reason for the relatively low ICC in M-GEDIS-UAV evaluation was found to be primarily attributable to disagreement among analysts in identifying those guidelines applicable to the target UAV interface. That is, while one analyst assigned a score for some criteria ("sufficient" or "not met"), another analyst might have considered the same design criteria to be irrelevant to the system interface. To understand whether the disagreement in applicable guidelines contributed to the low ICC, a follow-on assessment of the MP interface was conducted. The expert analysts reviewed all design criteria together and identified the ones that were relevant to the interface based on discussion. Subsequently, the expert analysts independently evaluated the interface in terms on the relevant design criteria. Using this approach, the ICC was found to increase to 0.834, indicating high reliability. These findings suggest that the low ICC in Experiment 1 was likely due to limited clarity on which guidelines were applicable to the interface, rather than whether a particular criterion was satisfied. The findings also suggest that human factors knowledge and familiarity with an interface are important for an accurate evaluation with the M-GEDIS-UAV. Some analyst training and collaboration may be necessary to ensure reliable interface evaluations.

4.2 Experiment 2: Further Assessment of Sensitivity of Modified GEDIS-UAV

The lack of sensitivity to the interface feature manipulation in Experiment 1 motivated further assessment of tool sensitivity for quantifying differences among interfaces. We introduced an additional "massive" UAV system information display into the experiment design. The "massive" display is another system information display alternative as part of the original MP interface. It presents a status of all system parameters in alphanumeric text (see Figure 5).

4.2.1 Interface Analysts and Evaluation Procedure. Due to higher inter-rater reliability among the expert analyst group in Experiment 1 results, only experts were used to evaluate the additional MP interface variation. As in Experiment 1, videos of the interface being used for the same two scenarios (mission control and alarm resolution) were presented to the experts. Based on the expert analysts' familiarity with the M-GEDIS-UAV and the interface features, it was determined that the massive data display manipulation would only impact five design indicator scores, including: Display Layout (DL), Information Presentation (IP), Color (C), Text (T), and Status and Devices (SD). Consequently, only these five indicators were re-evaluated by the expert analysts. The GEI for the massive data display was calculated using the five re-evaluated indicator scores and the other four original indicator scores. Subsequently, the GEIs for the massive data display were compared with the GEIs for the digital and analog display variations. It was expected that the massive data display would produce a significantly lower GEI than the other two interfaces due to further deviations of the design from M-GEDIS-UAV criteria.



```
roll            -0.023920  targetairspeed  0          my2       73     ch11in
pitch           0.000627   lowairspeed     False      mz2       -539   ch12in
yaw             353.9232   asratio         0          ax3       0      ch13in
groundcourse    0          groundspeed     0          ay3       0      ch14in
lat             -35.36326  ax              0          az3       0      ch15in
lng             149.1652   ay              0          gx3       0      ch16in
alt             0          az              -999       gy3       0      ch1out
altasl          1916.306   gx              2          gz3       0      ch2out
altoffsethome   0          gy              2          mx3       0      ch3out
gpsstatus       3          gz              2          my3       0      ch4out
gpshdop         1.21       mx              215        mz3       0      ch5out
satcount        10         my              73         failsafe  False  ch6out
lat2            0          mz              -539       rxrssi    0      ch7out
lng2            0          magfield        584.8718   ch1in     1500   ch8out
altasl2         0          accelsq         0.999      ch2in     1500   ch9out
gpsstatus2      0          gyrosq          3.464102   ch3in     1000   ch10out
gpshdop2        0          ax2             0          ch4in     1500   ch11out
satcount2       0          ay2             0          ch5in     1800   ch12out
groundspeed2    0          az2             -999       ch6in     1000   ch13out
groundcourse2   0          gx2             2          ch7in     1000   ch14out
altd1000        0          gy2             2          ch8in     1800   ch15out
altd100         0          gz2             2          ch9in     0      ch16out
airspeed        0          mx2             215        ch10in    0      ch3percent
```

Fig. 5. Massive Data Display (within the auxiliary box of MP interface).

Table 6. Evaluation Results Based on M-GEDIS-UAV

	GEI Mean (SD)	Tukey's HSD Grouping
Simple Digital Display	0.77 (0.08)	A
Analog Display	0.76 (0.06)	A
Massive Display	0.71 (0.07)	B

4.2.2 Results and Discussion. The aggregate GEIs for the three interface types are presented in Table 6. Due to normality violations for parametric testing, ranks of GEIs among analysts were submitted to the ANOVA procedure to yield a nonparametric test. Interface type was the only independent variable. Analyst served as a blocking factor in the statistical model.

Results revealed a highly significant effect of interface type ($F(2, 4) = 21.50, p = 0.007$) on the M-GEDIS-UAV scores. Tukey's HSD procedure was used for pairwise comparisons of the three interfaces. As shown in Table 5, treatment groups connected by the same letter are not significantly different. Results revealed the analog and simple digital display to produce significantly higher GEIs than the massive data display. The results demonstrated sensitivity of the M-GEDIS-UAV to interface manipulations.

To identify which of the five re-evaluated design indicators was most sensitive to the UAV control interface manipulation, a nonparametric ANOVA procedure was applied to each indicator score to test the significance of the display type manipulation. It was found that the "Text" indicator was significantly affected by the massive data display ($F(2, 4) = 8.57, p = 0.036$). A "hold-out" procedure was subsequently applied to the GEI calculation in which the "Text" indicator score was removed from the averaging of design indicator scores as part of the GEI. With this procedure, if the "display type" effect proved to be insignificant when the "Text" indicator was held out, then the result would indicate that the "Text" indicator was a major predictor of differences in the GEI scores for the interface variation. However, the "display type" effect remained highly significant ($F(2, 4) = 31, p = 0.004$), indicating that the "Text" was not the only predictor of the GEI score differences. Consequently, we applied the "hold-out" procedure to all possible combinations of the five re-evaluated design indicators for generating GEI scores. Results of this additional analysis revealed that the combination of all re-evaluated indicators was statistically important to sensitivity

Table 7. Evaluation Results Based on Original GEDIS-UAV

	GEI Mean (SD)	
	Raw	Scaled
Simple Digital Display	3.05 (0.80)	0.61 (0.16)
Analog Display	3.15 (0.70)	0.63 (0.14)
Massive Data Display	2.25 (0.70)	0.45 (0.14)

of the M-GEDIS-UAV to display manipulations; the display type manipulation was not significant in GEIs in the absence of the re-evaluated indicators ($F(2,4) = 1, p = 0.444$). Among all combinations of design indicators held-out from the GEI calculation, the subset of IP, T, SD accounted for the greatest degree of index variability. The display type manipulation was highly insignificant in GEIs absent of the identified indicators ($F(2,4) = 0.538, p = 0.621$). When each indicator was held-out separately from the analysis, the SD appeared to be the primary driver in significance of the display type manipulation in the GEIs. The removal of the SD score from the GEI led to an insignificant ANOVA result for display type ($F(2,4) = 6.5, p = 0.055$).

In summary, the additional sensitivity analysis indicated that changes in the M-GEDIS-UAV GEI score were driven by evaluation of multiple design indicators. In this case, the SD indicator appeared to be most sensitive to the manipulation as design variation focused on the style of system information presentation. It is important to note that the “Text" design indicator was also significantly influenced by display type. However, when included in the overall GEI calculation, this criterion appeared to be less important than the status and device display criteria. Although the original GEDIS-UAV GEIs showed a lower value for the massive data display, compared with the digital and analog displays, the difference was not significant.

4.3 Comparison Between Original and Modified GEDIS-UAV

4.3.1 Interface Analysts and Evaluation Procedure. To identify any improvements in interface evaluation with the M-GEDIS-UAV vs. the original GEDIS-UAV tool, we recruited a convenience sample of another three human factors experts to perform the same evaluation with the original tool. Similarly, all the analysts in this group completed at least two years of human-factors-related coursework, which allows them to understand the usability concepts presented in the original GEDIS-UAV. To avoid potential bias, the authors did not participate in the evaluation with the original GEDIS-UAV. The tool description and an example application of the GEDIS-UAV, as documented in the original publication [21], were used as training material for the analysts.

4.3.2 Results and Discussion. MP interface evaluation results based on the original GEDIS-UAV are presented in Table 7. Since the original GEDIS-UAV scores range from 0 to 5, they were rescaled to 0-1 to allow for comparison with M-GEDIS-UAV scores. A Spearman ranks correlation analysis was performed to identify any association between scores from the original and modified GEDIS-UAV tools. However, results did not reveal a significant correlation ($\rho = -0.03, p = 0.949$) between the responses, suggesting the two tools provided different evaluations. Subsequently, sensitivity and inter-rater reliability for the original GEDIS-UAV scores were compared with M-GEDIS-UAV results. Non-parametric analysis revealed no significant effect of interface type on the original GEDIS-UAV scores ($F(2,4) = 4.5, p = 0.095$), suggesting a lack of sensitivity to interface manipulations. Regarding inter-rater reliability, the original GEDIS-UAV indicator scores for the simple digital interface were used for ICC calculation. Results revealed low inter-rater reliability (ICC = 0.235) and an insignificant correlation ($F(9,15.6) = 2.47, p = 0.06$) among human factors expert scores. These findings suggest that the M-GEDIS-UAV presented improvement in terms of sensitivity to interface manipulation and inter-rater reliability, as compared to the original GEDIS-UAV.

5 CONCLUSIONS

5.1 Summary

Supervisory control interfaces are necessary for UAV safe operation and optimal system performance. Interface evaluation tools can help identify design deviations from principles and guidelines early in the design process and provide means for interface improvement to ultimately prevent damage or loss of vehicles. In this study, we reviewed existing evaluation tools for UAV supervisory control interfaces and identified some issues with various methods. Using the GEDIS-UAV evaluation tool as a basis [21], we developed the M-GEDIS-UAV with enhanced design criteria content and organization as well as a revised interface scoring approach.

The new tool provides a number of improvements over the original GEDIS-UAV. First, the design indicators provide a framework for understanding and assessing UAV interfaces. This approach ensures that the interface evaluation follows a systematic procedure and does not overlook any major aspects of design. Moreover, the sub-indicators represent a comprehensive set of usability principles and functionality features and characteristics. All evaluation criteria are supported by literature and provide direct guidance on interface redesign. Although some human factors training is required for understanding the sub-indicator criteria, the comprehensive list reduces the workload of an analyst by eliminating the need to look-up references. Another improvement is that the M-GEDIS-UAV minimizes subjectivity in interface evaluations. By comparing an interface with concrete design guidelines and determining the degree of conformance, there is a reduction in response biases [11] that are common in traditional rating-based approaches. Last but not least, the M-GEDIS-UAV can be applied across UAV interfaces providing different functionality. When dealing with less complicated UAV interfaces, the analyst can easily adjust the tool by excluding non-applicable sub-indicators from the evaluation. Similar to the original tool, the M-GEDIS-UAV allows the calculation of an evaluation score. This allows quantitative comparison among various interfaces. With additional research, it is also possible to determine a score recommendation for UAV interface design.

The M-GEDIS-UAV was applied to several interfaces to assess validity and sensitivity for interface evaluation. In general, the tool produced consistent scoring outcomes for a specific interface design across multiple expert analysts. Novice analysts demonstrated relatively lower consistency in scoring and would likely benefit from additional training on the method, as well as a target interface, for reliable outcomes. Additionally, identification of relevant design criteria for an interface evaluation should be based on expert group discussion to ensure reliability in tool outcomes across experts. The tool also proved sensitive to interface design variations involving changes in multiple design indicators/features including style and information presentation.

5.2 Limitations and Future Work

There are a couple of limitations associated with the present study. Although the main objective of this research was to develop the M-GEDIS-UAV tool for assisting UAV control interface designers, in the tool validation, only a small number of analysts applied the method to different interfaces. A larger sample of analysts could benefit the investigation in terms of inter-rater reliability and gathering thorough feedback on tool application. In terms of analyst training, future research should identify eligibility of analysts and investigate appropriate training for novices. Providing detailed explanation of terminology used as part of criteria might also be beneficial to guide novice analysts in understanding design guidelines. Moreover, the tool has only been fully applied to one

UAV control interface. Additional application testing and feedback analysis should be conducted to further generalize the utility of the M-GEDIS-UAV. Furthermore, future study should further assess the utility of M-GEDIS-UAV outcomes (GEIs and indicator scores) for predicting UAV operator performance and cognitive workload in UAV operations. Scatterplots on GEIs and performance measures could reveal "knee-points" in operator performance (i.e., significant decreases or increases) in association with certain levels of GEI. These knee-points could ultimately be identified as criteria for determining whether interface designs are "acceptable." Such criteria could be helpful for supporting the UAV control interface design industry. Currently, full evaluations with the M-GEDIS-UAV can last 2 hours for an expert analyst. The identification of relevant design criteria for interface evaluation may take an additional 1–2 hours. To further simplify the tool, future research could identify critical design criteria within each indicator that serve as "Go" or "No-Go" criteria in the conceptual or detailed phases of a design process. That is, an interface must at least comply with such critical criteria; otherwise, redesign of the interface would be mandatory. A hierarchical collection of design criteria could be used to identify for an analyst the most important interface elements for re-design and might reduce evaluation time. Finally, in this study, we demonstrated the M-GEDIS-UAV as a "reactive" tool for assessing existing UAV interface designs and identifying issues. However, the tool may also be applied proactively through application to early ("paper-and-pencil") interface prototypes, as a basis for guiding a design process.

REFERENCE

[1] Daniel Afergan, Evan M. Peck, Erin T. Solovey, Andrew Jenkins, Samuel W. Hincks, Eli T. Brown, Remco Chang, and Robert J. K. Jacob. 2014. Dynamic difficulty using brain metrics of workload. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*. 3797–3806. Retrieved from http://dl.acm.org/citation.cfm?id=2557230.

[2] J. Y. C. Chen, M. J. Barnes, and M. Harper-Sciarini. 2011. Supervisory control of multiple robots: Human-performance issues and user-interface design. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* 41, 4 (2011) 435–454. DOI: https://doi.org/10.1109/TSMCC.2010.2056682

[3] M. L. Cummings, A. Kirschbaum, A. Sulmistras, and J. T. Platts. 2006. STANAG 4586 human supervisory control implications. In *Proceedings of the Unmanned Vehicle Systems Canada Conference*.

[4] M. L. Cummings, Kevin Myers, and Stacey D. Scott. 2006. Modified cooper harper evaluation tool for unmanned vehicle displays. In *Proceedings of UVS Canada: Conference on Unmanned Vehicle Systems Canada*. Retrieved from http://www.eng.uwaterloo.ca/~s9scott/wiki/uploads/Main/cummings_uvs2006.pdf.

[5] Henrica C. W. De Vet, Caroline B. Terwee, Lidwine B. Mokkink, and Dirk L. Knol. 2011. *Measurement in Medicine: A Practical Guide*. Cambridge University Press.

[6] Alan Dix. 2009. *Human-computer Interaction*. Springer. Retrieved from http://link.springer.com/10.1007/978-0-387-39940-9_192.

[7] Birsen Donmez, M. L. Cummings, Hudson D. Graham, and Amy S. Brzezinski. 2010. Modified cooper harper scales for assessing unmanned vehicle displays. In *Proceedings of the 10th Performance Metrics for Intelligent Systems Workshop*. 235–242. Retrieved from http://dl.acm.org/citation.cfm?id=2377620.

[8] Paul Fahlstrom and Thomas Gleason. 2012. *Introduction to UAV Systems*. John Wiley & Sons. Retrieved from https://books.google.com/books?hl=en&lr=&id=uLsNtm99IWYC&oi=fnd&pg=PT9&dq=Introduction+to+UAV+systems&ots=lPKTzbaXLA&sig=L2tkHoePRw_UJh0AWS-NIYBsjtI.

[9] Terrence Fong and Charles Thorpe. 2001. Vehicle teleoperation interfaces. *Auton. Robots* 11, 1 (2001), 9–18.

[10] Christian Fuchs, Clark Borst, Guido C. H. E. de Croon, M. M. René van Paassen, and Max Mulder. 2014. An ecological approach to the supervisory control of UAV swarms. *Int. J. Micro Air Veh.* 6, 4 (2014), 211–229.

[11] Adrian Furnham. 1986. Response bias, social desirability and dissimulation. *Personal. Individ. Differ.* 7, 3 (1986), 385–400.

[12] R. P. Harper and G. E. Cooper. 1986. Handling qualities and pilot evaluation. *J. Guid. Control Dyn.* 9, 5 (1986), 515–529. DOI: https://doi.org/10.2514/3.20142

[13] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology*, Peter A. Hancock and Najmedin Meshkati (Ed.). North-Holland, 139–183. Retrieved from http://www.sciencedirect.com/science/article/pii/S0166411508623869.

[14] Erik Hollnagel. 2003. *Handbook of Cognitive Task Design*. CRC Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=dElPH0ruR-sC&oi=fnd&pg=PP1&dq=Handbook+of+Cognitive+Task+Design&ots=EGmVq44LDG&sig=JDYSbTWmGljZsYB30icGrCtQC5E.

[15] Curtis M. Humphrey and Julie A. Adams. 2011. Analysis of complex team-based systems: Augmentations to goal-directed task analysis and cognitive work analysis. *Theor. Issues Ergon. Sci.* 12, 2 (2011), 149–175.

[16] Javier Irizarry, Masoud Gheisari, and Bruce N. Walker. 2012. Usability assessment of drone technology as safety inspection tools. *J. Inf. Technol. Constr.* 17 (2012), 194–212.

[17] Patrick W. Jordan, B. Thomas, Ian Lyall McClelland, and Bernard Weerdmeester. 1996. *Usability Evaluation In Industry*. CRC Press.

[18] David B. Kaber, Jennifer M. Riley, and Kheng-Wooi Tan. 2002. Improved usability of aviation automation through direct manipulation and graphical user interface design. *Int. J. Aviat. Psychol.* 12, 2 (2002), 153–178.

[19] David B. Kaber, Noa Segall, Rebecca S. Green, K. Entzian, and S. Junginger. 2006. Using multiple cognitive task analysis methods for supervisory control interface design in high-throughput biological screening processes. *Cogn. Technol. Work* 8, 4 (2006), 237–252.

[20] Amy Kransteuber, Jeffrey Wilbert, Mike Castaneda, and Erin Alves. 2012. *Unmanned Aircraft Systems Ground Control Station Human-Machine Interface: Development & Standardization Guide*. Office of the Under Secretary of Defense.

[21] Salvador Lorite, Adolfo Muñoz, Josep Tornero, Pere Ponsa, and Enric Pastor. 2013. Supervisory control interface design for unmanned aerial vehicles through GEDIS-UAV. In *Proceedings of the International Conference on Human-Computer Interaction*. 231–240. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-39232-0_26.

[22] Jen-Li Lu, Ruey-Yun Horng, and Ching-Jung Chao. 2013. Design and test of a situation-augmented display for an unmanned aerial vehicle monitoring task 1. *Percept. Mot. Skills* 117, 1 (2013). Retrieved from http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=00315125&AN=92036019&h=cK9JOtrvTO6ijRgAYVGJAPUh3ytj0RVluePTWRZ6Abt6O85DdttqGA0wLDq%2FvlkPxwQg02QJ8wO%2FF5Qyn1bRsw%3D%3D&crl=c.

[23] Kenneth O. McGraw and Seok P. Wong. 1996. Forming inferences about some intraclass correlation coefficients. *Psychol. Methods* 1, 1 (1996), 30.

[24] Justin Menda, James T. Hing, Hasan Ayaz, Patricia A. Shewokis, Kurtulus Izzetoglu, Banu Onaral, and Paul Oh. 2011. Optical brain imaging to enhance UAV operator training, evaluation, and interface development. *J. Intell. Robot. Syst.* 61, 1–4 (2011), 423–443.

[25] Jakob Nielsen. 1994. *Heuristic Evaluation in Usability Inspection Methods (RL Mack)*. John Wiley & Sons, Inc., New York.

[26] Jakob Nielsen. 1996. Usability metrics: Tracking interface improvements. *IEEE Softw. Los Alamitos* 13, 6 (Nov. 1996), 12–13. DOI: http://dx.doi.org/10.1109/MS.1996.10031

[27] Jakob Nielsen and Rolf Molich. 1990. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 249–256. Retrieved from http://dl.acm.org/citation.cfm?id=97281.

[28] Donald A. Norman. 2013. *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books.

[29] John O'Hara, W. Brown, P. M. Lewis, and J. J. Persensky. 2002. Human-system interface design review guidelines (NUREG-0700, Rev 2). *Wash. D.C. U.S. Nucl. Regul. Comm.* (2002), 12.

[30] Pere Ponsa Asensio, Beatriz Amante García, and Marta Díaz Boladeras. 2010. Ergonomic design applied in a sugar mill interface. *Lat. Am. Appl. Res.* 40, 1 (2010), 27–33.

[31] Pere Ponsa and Marta Díaz. 2007. Creation of an ergonomic guideline for supervisory control interface design. In *Proceedings of the International Conference on Engineering Psychology and Cognitive Ergonomics*. 137–146. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-73331-7_15.

[32] Leslie G. Portney and Mary P. Watkins. 2015. *Foundations of Clinical Research: Applications to Practice*. F. A. Davis, Philadelphia, PA.

[33] Jan Maarten Schraagen, Susan F. Chipman, and Valerie L. Shalin. 2000. *Cognitive Task Analysis*. Psychology Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=gyt5AgAAQBAJ&oi=fnd&pg=PP1&dq=Cognitive+Task+Analysis&ots=cfI2wxN5ta&sig=L9Lo_HzsdCO9sfuId5MtjrfT5p8.

[34] Ben Shneiderman and Catherine Plaisant. 1987. Designing the user interface: Strategies for effective human-computer interaction. *ACM SIGBIO Newslett.* 9, 1 (1987), 6.

[35] Kevin W. Williams. 2004. A Summary of Unmanned Aircraft Accident/Incident Data: Human Factors Implications. DTIC Document. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA460102.

[36] Veysel Yesilbas and T. Steven Cotter. 2014. Structural analysis of hfacs in unmanned and manned air vehicles. In *Proceedings of the International Annual Conference of the American Society for Engineering Management*. 1–7. Retrieved from http://search.proquest.com/docview/1705171380/abstract/5BCAB252E0054797PQ/1.

[37] Wenjuan Zhang, David Feltner, James Shirley, and David Kaber. 2016. Unmanned aerial vehicle control interface design and cognitive workload: A constrained review and research framework. DOI: 10.1109/SMC.2016.7844502

[38] RTCA, Inc. 1992. DO-178B, "Software Considerations in Airborne Systems and Equipment Certification," Oct. '98. Retrieved from http://www.dcs.gla.ac.uk/~johnson/teaching/safety/reports/schad.html.

[39] Department of Defense. 1998. MIL-STD-1472F design criteria standard: Human engineering. Retrieved from http://segoldmine.ppi-int.com/documents/mil-std-1472d-notice-1-military-standard-human-engineering-design-criteria-military-system.

[40] ISO 11064-7. 2006. Ergonomic design of control centers, *Part 7: Principles for the evaluation of control centres*. ISO. Retrieved from http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=22470.

[41] MATB-II. 2019. Retrieved from https://matb.larc.nasa.gov/.

Received January 2018; revised September 2019; accepted October 2019