



# Assessing Operator Psychological States and Performance in UAS Operations

Jinchao Lin<sup>(✉)</sup>, Gerald Matthews, Lauren Reinerman-Jones,  
and Ryan Wohleber

Institute for Simulation and Training (IST), University of Central Florida (UCF),  
Orlando, FL, USA

{jlin, gmatthew, lreiner, rwohlebe}@ist.ucf.edu

**Abstract.** Assessment for understanding, predicting, and improving human performance and system design is a key for human-computer interaction (HCI) research. Assessments can be behavioral, physiological, performance-based, and phenomenological. Assessments are important in a variety of domains, including unmanned vehicle operations, human-robot teaming, nuclear power plant operations, etc. This paper will discuss assessment approaches in the domain of unmanned aerial systems (UAS) operations to identify and quantify explanatory constructs, such as psychological states, workload, and performance. It will also discuss implications for evaluating improvements in human performance in UAS operations. Specifically, this paper will examine metrics that can be utilized to gauge the impact of demand factors on workload, task performance, operator dependence on automation, and stress response.

**Keywords:** UAS · Workload · Stress · Performance · Assessment

## 1 Introduction

Nowadays, machines with advanced technology are capable of performing varieties of jobs which are currently performed by humans. Although some machines are designed to do their jobs automatically, human operators, at least in a supervisory role, are still required in many cases. The development of technology makes complex human-machine systems possible and brings numerous benefits, but it may also lead to human factors challenges in terms of operator performance. Assessing operators' psychological states and performance is essential for understanding, predicting, and improving human performance and system design in those complex human-machine systems. This paper will take a simulation study of unmanned aerial systems (UAS) operations as an example to discuss assessment approaches for identifying and quantifying explanatory constructs. It focuses especially on the challenges of measuring various facets of overload, such as psychological states, subjective workload, and performance impairment.

In current UAS mission configurations, operators are often more at risk from underload than overload, due to requirements to loiter for extended periods until a target appears. However, workload issues may change as future UAS will allow multi-aircraft control (MAC) by a single operator. This is anticipated to be a

particularly time-critical, and cognitively demanding, form of multi-tasking work [1, 2]. Automation of system functions is essential for MAC, but the necessity of keeping the operator in the loop and capable of coordinating multiple vehicles limits the extent to which automation can mitigate the increased workload associated with multi-tasking. Evaluating system design and operator competencies requires assessment methodologies that can detect various expressions of overload, including excessive subjective workload, stress, and performance impairment. One of the least understood metrics for overload is change in trust [3]. Appropriate trust in automation may be especially important in overload situations, but there is rather little evidence on the extent to which trust measures converge or diverge with those for operator overload.

### 1.1 Workload and Stress in UAS Operation

UAS operations often involve considerable task demand variation which may be accompanied by both stress, and changes in operator performance. In Hancock and Warm's [4] theoretical model for stress and performance, individuals can adapt effectively to some levels of task demands without showing significant performance decrement. However, both extreme overload and underload could result in failures in such adaptation. In terms of MAC, overload from sources including monitoring displays, flight control, navigation, communication, and mission management is the primary challenge [5]. As the number of vehicles controlled increases, operators may become more vulnerable to overload due to the limited resources taxed by multiple demand factors [6].

Automation plays a critical role in mitigating the workload associated with MAC by keeping cognitive demands to a manageable level [5]. Calhoun et al. [1] showed that automation benefits for performance transferred from the subtask automated to additional subtasks, implying that automation may free some general attentional resources. However, automation may fail to mitigate workload if it is poorly designed or used inappropriately [7].

The impact of automation on workload may also depend on the level of automation (LOA), i.e., the extent of the tradeoff between operator control and delegation of control to the machine [8]. Generally, higher LOAs reduce operator workload, but may impair vigilance and situation awareness. MAC studies typically envisage intermediate levels of control such that the operator must check and possibly over-ride the recommendations of the automation. Effects of LOA on performance and on trust in automation are mixed in empirical studies: the optimal LOA may depend on task demands and other aspects of system configuration [1, 9].

Stress commonly accompanies high workload [10]. Indeed, UAS operators report that task demand factors such as interface difficulties and inefficiencies in control procedures contribute to stress, along with occupational health factors [11]. However, stress may be less than in conventional flying. Skilled performers are often able to mitigate task-induced stress by developing strategies that prevent catastrophic performance failures and strategies for emotion-regulation [12].

In an empirical study using a MAC simulation, Wohleber et al. [13] examined the impact of task demand on both subjective and physiological stress indices. High task demand produced substantial increases in workload and distress, but only affected

certain stress indices. Specifically, demand increased high-frequency activity in the electroencephalogram (EEG), but did not affect cardiac response. These results suggested that high demand produced a relatively subtle, “cognitive” form of stress, rather than the classic fight-or-flight response, which would have elevated heart rate [13].

The present study focused on assessing multiple impacts of task demands during MAC operations. The simulation required multi-tasking with support from automation. It was configured to prioritize performance of surveillance activities common in UAS operations. Sustained monitoring often imposes high workload leading to depletion of processing resources and vigilance decrement [14], as well as subjective distress and loss of task engagement [15]. The study evaluated the extent to which the different indicators of overload converged under high task demands, as well as investigating the overload effect on dependence on automation.

## 1.2 Assessment of Operator Response to Cognitive Demands

Strain on UAS operators takes a variety of forms, influencing multiple objective and subjective responses. In one of the ISR tasks used in the present experiment, participants were asked to monitor displays to identify targets and take actions based on the discrimination of enemy and friendly tanks. In this case, accuracy, which is the percentage of correct actions, can be a straightforward assessment reflecting operator performance. However, the multi-component nature of the task also suggests that demand factors might influence subtask prioritization and possible neglect of subtasks [16]. Automated systems also raise the issue of choosing metrics for trust and reliance on the automation in decision-making. Trust is typically seen as a quality of operator state antecedent to behavior [3]. Behavioral indices seek to capture the different ways in which operator use of automation is suboptimal [17]. In the context of automated alarms, performance is typically assessed in terms of two measures [18]: compliance (taking action following the alarm) and reliance (taking no action in the absence of an alarm). However, this distinction may not be applicable to routine binary stimulus discrimination, where neither option constitutes an alarm requiring urgent response. Therefore, in this study, the metric of dependence on automation, the extent to which the operator follows the recommendations of the automation across all trials [19], was adopted as a behavior index for trust.

Workload and stress can be assessed using objective, physiological response and subjective scales [20]. The current study used the subjective measures sensitive to task demands in previous studies of simulated UAS [21] and unmanned ground vehicle (UGV) [22] operations: the NASA-TLX [23] workload measure, and the Dundee Stress State Questionnaire that assesses affective, motivational and cognitive aspects of subjective state (DSSQ) [10]. The DSSQ represents a multidimensional perspective on stress: task demands and environmental pressures elicit a range of qualitatively different subjective state responses [24].

The present research focused on the three higher-order DSSQ factors extracted from the primary dimensions: distress, task engagement, and worry. Distress is driven by a sense of being overloaded and lack of control over the task environment, coupled with negative affect [25]. Task demand manipulations often influence distress and workload concurrently [24, 25], and the distress scale correlates with NASA-TLX

workload [10]. In unmanned vehicle studies, multi-tasking demands elevate distress substantially [17, 22]. Task engagement reflects energy, task motivation, and alertness. Operators managing UAS may be prone to both upwards and downwards shifts in task engagement. For example, complex, challenging tasks including game-like scenarios involving unmanned vehicle control tend to elevate task engagement [2, 25]. In a UGV study, Abich et al. [22] found that both multi-tasking and higher event rates produced moderate elevations of engagement. On the other hand, prolonged, monotonous tasks such as vigilance typically lower task engagement substantially [24, 26]. Guznov et al. [27] found moderate-magnitude engagement decline during a 30-min UAS surveillance mission, consistent with reports of monotony and fatigue during real-life missions of substantially longer durations [28]. The third higher-order DSSQ dimension, worry, corresponds to self-focused attention, low self-esteem, and high cognitive interference. In unmanned vehicle studies, it tends to be less sensitive to task demand manipulations than distress and task engagement [22]. Poorly designed automation that threatens the operator's sense of personal competence might elevate worry [24]. Worry also overlaps with mind-wandering which may develop during prolonged operations [28].

### 1.3 Aims and Hypotheses

The study aimed to apply a multivariate assessment strategy [20] to profile the workload, stress and performance changes associated with high task demands during simulated UAS operation. We tested hypotheses derived from the Standard Capacity Model (SCM) [12]. The model captures the typical assumption of workload researchers that overload is associated with an insufficiency of general attentional resources to meet demands for processing. Lack of resources drives objective performance impairment and high subjective workload and distress. The effects of resource insufficiency on trust and automation-dependence have not been much researched; however, Parasuraman and Manzey [29] theorized that high task demands would tend to increase dependence on automation as a compensatory strategy (depending also on the configuration of the automation). In the limiting case, the different responses to overload would be interchangeable as indices of a general overload or resource insufficiency syndrome. However, we expected that we would in fact find divergences between different measures as in previous, related studies [21, 22, 30]. We assessed the strength of impact of task demands on dependent variables in terms of Cohen's effect size measure  $d$ . Conventionally,  $d$  values of .2, .5, and .8 correspond to small, medium and large effect sizes. The study also manipulated LOA. In the present study, LOA was of interest primarily as a potential moderator of task demand effects, but we report its effects briefly.

## 2 Method

### 2.1 Participants and Experimental Design

A total of 101 college students (42 men, 59 women,  $M_{\text{age}} = 18.95$ ,  $SD = 1.80$ ) participated this study for course credit. Participants were healthy individuals between 18

and 40 years old representing the age group and educational level of the enlisted military service core that may be recruited for future UAS operations. Participants who may be vulnerable to adverse reactions, such as excessive stress, resulting from the test environment were excluded. All participants reported having normal or corrected to normal vision, color vision, normal hearing, and English fluency.

A 2 (task demand: high versus low)  $\times$  2 (LOA: management-by-consent versus management-by-exception) between-subjects factorial design was adopted in this study to assess demand factor impacts and operator performance. Twenty-six participants took part in the low task demand/management-by-exception condition; there were twenty-five participants in the other three conditions.

## 2.2 Simulation

The ALOA (Adaptive Levels of Autonomy) multi-UAS research test bed developed by OR Concepts Applied [1] was used for the study. This simulation supports task manipulations representing UAS operations in needed complexity and realism (see Fig. 1). Nine tasks, including target allocation and rerouting, two surveillance tasks, aircraft identification, status monitoring, decision making, and information retrieval, were designed to represent the task demands for a single operator managing a fleet of four aircrafts with an automation aid at the same time. The LOAs were varied in two intermediate levels with high reliability (correct 80% of the time) to support the two primary surveillance tasks, Image Analysis and Weapon Release authorization. Management-by-consent required participants to accept or change the option recommended by the automation. Alternatively, with management-by-exception, the system was set to act on the option recommended by the automation automatically unless a different option was selected before the availability of operator response was timed out (30 or 20 s based on tasks).

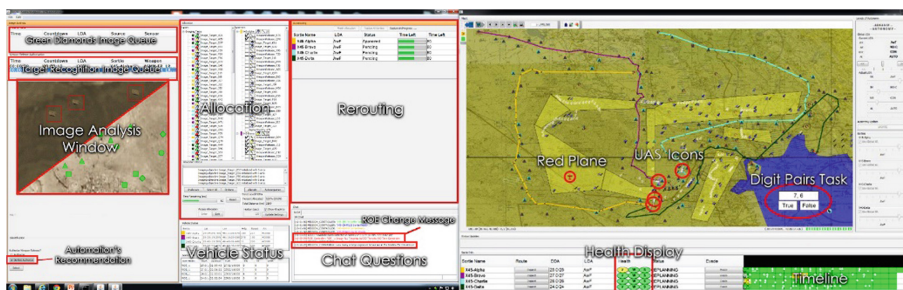


Fig. 1. Task interface for MAC operation in the ALOA UAS simulation

Task frequencies of the secondary tasks were manipulated to create task demand variation across conditions in one hour long experimental scenarios whereas task frequencies of two primary surveillance tasks were held constant. There were 6 tasks or 14 tasks per minute to induce low and high task demand respectively. Most secondary tasks required responses to visual or audio signals, searching and retrieving

information, or comparing digit pairs. Image Analysis and Weapon Release authorization tasks were timed for 30 and 20 s respectively. The task response would be recorded as a “miss” if there is no operator response before task availability was timed out.

In the Image Analysis task, images with an overlay of 19–26 green symbols varying in shapes, including diamonds, squares, circles, and triangles, were shown in the task panel. Participants were asked to count the number of diamonds and select the number from eight options. The automated aid system recommended one from the eight options by highlighting it. The reliability of the automation was set to be 80% correct.

In the Weapon Release authorization task, participants were asked to distinguish hostile tanks from allied tanks and detect whether the hostile tanks in given picture were correctly marked. The tanks differed in body width and barrel length subtly. The pictures were degraded in quality to increase the difficulty in discrimination. The automation aid system recommended one option from “authorize” or “do not authorize”. Also, reliability was set to be 80% correct.

### 2.3 Subjective Measures

*Demographics Questionnaire.* We developed a 21-item demographics questionnaire with questions on a range of biographical information, including age, gender, health status, education level, computer expertise, and gaming experience and expertise.

*Stress state.* The short, 21-item version of the Dundee Stress State Questionnaire (DSSQ) [25] was used to measure three higher order dimensions of subjective states in terms of task engagement, distress, and worry. Participants rated the accuracy of statements about their subjective states using a 5-point Likert.

*Workload.* The NASA-Task Load Index [23] consists of six 0–100 rating scales, including mental demand, physical demand, temporal demand, performance, effort, and frustration. Overall workload was calculated as the unweighted mean of the six ratings.

### 2.4 Performance-Based Measures

Three performance metrics for the two high priority surveillance tasks, Image Analysis and Weapon Release authorization, were analyzed. Accuracy was defined as the percentage of correct responses. Dependence on automation was defined as the percentage of trials on which the participant followed the recommendation from the automation. Neglect was defined as the frequency of items that appeared in the task window but were not opened by the participant. Detailed performance metric formulas for Image Analysis and Weapon Release authorization tasks are listed in Tables 1 and 2. Correct responses were labeled as “hit” (agree with automation) and “correct rejection” (disagree with automation). Incorrect answers were labeled as “near/far miss” (agree with automation) and “false alarm” (disagree with automation). In low LOA condition, if no action was taken before time-out, the task was recorded as a “true miss”.

**Table 1.** Performance metrics in the Image Analysis task

	Formula
<b>Low LOA</b>	
Accuracy	$\frac{Hit + CorrectRejection}{Hit + CorrectRejection + NearMiss + FarMiss + FalseAlarm + TrueMiss} \times 100\%$
Dependence on Automation	$\frac{Hit + NearMiss + FarMiss}{Hit + CorrectRejection + NearMiss + FarMiss + FalseAlarm + TrueMiss} \times 100\%$
Neglect	Number of tasks which the participant never opened
<b>High LOA</b>	
Accuracy	$\frac{Hit + CorrectRejection}{Hit + CorrectRejection + NearMiss + FarMiss + FalseAlarm} \times 100\%$
Dependence on Automation	$\frac{Hit + NearMiss}{Hit + CorrectRejection + NearMiss + FarMiss + FalseAlarm} \times 100\%$
Neglect	Number of tasks which the participant never opened

**Table 2.** Performance metrics in the Weapon Release authorization task

	Formula
<b>Low LOA</b>	
Accuracy	$\frac{Hit + CorrectRejection}{Hit + CorrectRejection + NearMiss + FalseAlarm + TrueMiss} \times 100\%$
Dependence on Automation	$\frac{Hit + NearMiss}{Hit + CorrectRejection + NearMiss + FalseAlarm + TrueMiss} \times 100\%$
Neglect	Number of tasks which the participant never opened
<b>High LOA</b>	
Accuracy	$\frac{Hit + CorrectRejection}{Hit + CorrectRejection + NearMiss + FalseAlarm} \times 100\%$
Dependence on Automation	$\frac{Hit + NearMiss}{Hit + CorrectRejection + NearMiss + FalseAlarm} \times 100\%$
Neglect	Number of tasks which the participant never opened

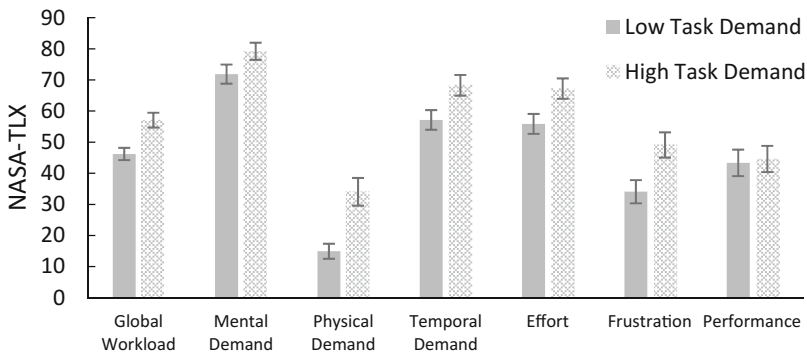
## 2.5 Procedure

Following an informed consent procedure, participants were instructed to complete the pre-task survey set, including the Demographic Questionnaire, and the pre-task DSSQ. After completing pre-task surveys, training started with an introduction using Power-Point slides, followed by a live simulation demonstration and hands-on practice. A “cheat sheet” about all the tasks was provided for quick reference. Training took approximately 60 min. Participants have to be qualified in the practice. Before the experimental task, researcher repeated instructions for simulation controls briefly and emphasized task priorities. The task ran for 60 min. Finally, participants were instructed to complete the post-task DSSQ and NASA-TLX, prior to debriefing. All the sessions in total were completed within three hours.

## 3 Results

### 3.1 Workload

Bonferroni-corrected *t*-tests were run to test the effects of experimental manipulations. It was confirmed that workload (NASA-TLX global workload) was significantly higher in high task demand conditions ( $M = 57.1$ ) than in low task demand conditions ( $M = 46.2$ ),  $t(99) = -3.52$ ,  $p = .001$ ,  $d = .70$ . According to NASA-TLX, the manipulation of task demand successfully elicited higher workload in all aspects, including mental demand,  $t(99) = -1.78$ ,  $p = .079$ ,  $d = .35$ ; physical demand,  $t(99) = -3.77$ ,  $p < .01$ ,  $d = .75$ ; temporal demand,  $t(99) = -2.43$ ,  $p < .05$ ,  $d = .48$ ; effort,  $t(99) = -2.47$ ,  $p < .05$ ,  $d = .49$ , and frustration,  $t(99) = -2.73$ ,  $p < .01$ ,  $d = .54$ , in high task demand conditions (see Fig. 2). However, there was no difference in self-reported performance,  $t(99) = -.21$ ,  $p = .835$ .



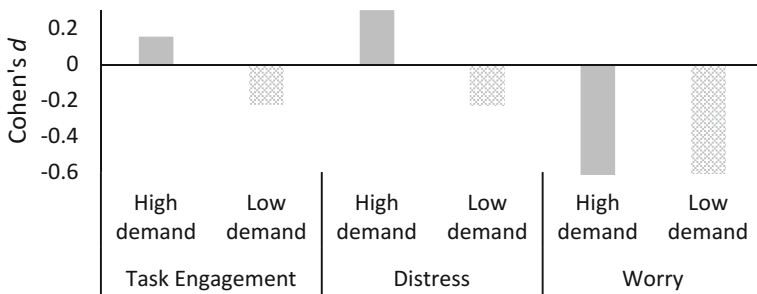
**Fig. 2.** NASA-TLX workload factor ratings in low/high task demand conditions. Error bars represent standard errors.

Bonferroni-corrected *t*-tests were also computed to check the impact of LOA manipulations. No significant self-rated workload differences were found between different LOA conditions.



### 3.2 Stress State

A series of  $2 \times 2 \times 2$  (LOA  $\times$  task demand  $\times$  pre- vs. post-task) mixed-model ANOVAs were run for each stress state factors, including task engagement, distress, and worry, to test the effects of experimental manipulations on subjective states. The effect sizes (Cohen's  $d$ ) for the stress state changes in different task demand conditions are shown in Fig. 3. The results from ANOVA for task engagement showed a near significant interaction between pre-/post-task and task demand,  $F(1, 97) = 3.65$ ,  $p = .059$ ,  $\eta_p^2 = .04$ . Pre-task engagement levels were similar in low demand ( $M = 21.75$ ) and high demand ( $M = 21.32$ ) groups, but following task performance engagement decreased under low demand ( $M = 20.29$ ,  $d = -.22$ ) but increased slightly under high demand ( $M = 22.16$ ,  $d = .16$ ). There was another significant interaction between pre-/post-task and task demand for distress,  $F(1, 97) = 7.81$ ,  $p < .01$ ,  $\eta_p^2 = .07$ . In the high task demand condition, participants reported greater distress ( $M = 10.80$ ) after task exposure, compared to the pre-task baseline ( $M = 9.22$ ,  $d = .30$ ). In the low task demand condition, participants were less distressed after task exposure ( $M = 6.90$ ,  $d = -.23$ ) than the pre-task baseline ( $M = 8.14$ ). Regarding worry, a significant main effect for pre-/post-task was found,  $F(1, 97) = 46.14$ ,  $p < .01$ ,  $\eta_p^2 = .32$ . Worry decreased in all conditions. In low task demand condition, worry levels decreased after task exposure ( $M = 10.20$ ), compared to the baseline ( $M = 12.92$ ); in high task demand condition, worry levels decreased after task exposure ( $M = 10.34$ ), compared to the baseline ( $M = 14.18$ ). The decrease of worry was greater in high task demand condition ( $d = -.70$ ) than in low task demand condition ( $d = -.60$ ).



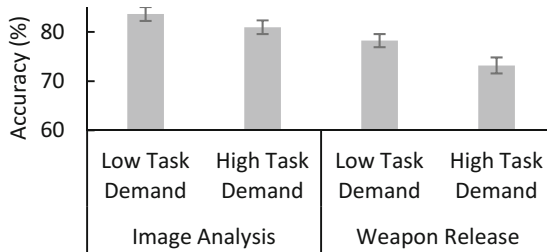
**Fig. 3.** Effect sizes for stress state changes (pre-task to post-task) in low/high task demand conditions.

### 3.3 Task Performance

A series of  $2 \times 2 \times 2$  (LOA  $\times$  task demand  $\times$  task type) mixed-model ANOVAs were computed to assess the UAS operation performance, in terms of accuracy, dependence on automation, and neglect, under different demand factors.

### Accuracy

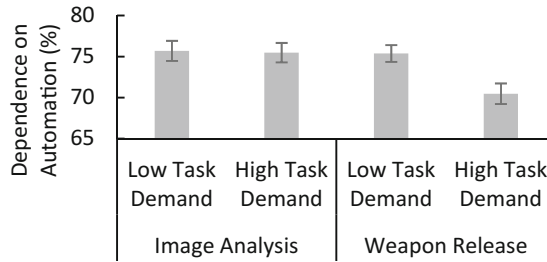
Participants performed less accurately in the Weapon Release authorization task ( $M = 75.7$ ) than in the Image Analysis task ( $M = 82.3$ ),  $F(1, 91) = 23.91$ ,  $p < .01$ ,  $\eta_p^2 = .21$  (see Fig. 4). Another main effect of task demand was also significant for accuracy,  $F(1, 91) = 5.87$ ,  $p < .05$ ,  $\eta_p^2 = .06$ . Participants in low task demand groups ( $M = 80.9$ ) achieved greater accuracy than those in high task demand groups ( $M = 77.1$ ) in the surveillance tasks. Accuracy in Weapon Release authorization task ( $d = -.49$ ) seemed to be more vulnerable to high task demand than Image Analysis task ( $d = -.28$ ), even though the interaction between task type and task demand was not significant.



**Fig. 4.** Task performance (accuracy) in the Image Analysis and the Weapon Release authorization tasks for different task demand conditions. Error bars represent standard errors.

### Dependence on Automation

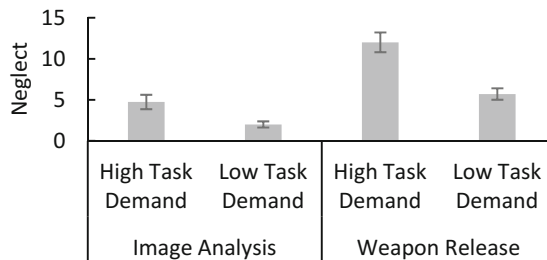
Dependence on automation was greater in the Image Analysis task ( $M = 75.6$ ) than in the Weapon Release authorization task ( $M = 72.9$ ),  $F(1, 91) = 5.91$ ,  $p < .05$ ,  $\eta_p^2 = .06$ . Result revealed a significant main effect of LOA for dependence,  $F(1, 91) = 5.11$ ,  $p < .05$ ,  $\eta_p^2 = .05$ . High LOA groups ( $M = 75.64$ ) were more dependent on automation than low LOA groups ( $M = 72.76$ ). LOA had a stronger effect on dependence on automation in Weapon Release authorization task ( $d = .46$ ) than in Image Analysis task ( $d = .25$ ). Also, a near significant main effect of task demand for dependence on automation was found,  $F(1, 91) = 3.92$ ,  $p = .051$ ,  $\eta_p^2 = .04$  (see Fig. 5). Participants showed greater dependence on automation in low task demand conditions than in high task demand conditions. In addition, the interaction between task type and task demand was also significant,  $F(1, 91) = 4.76$ ,  $p < .05$ ,  $\eta_p^2 = .05$ . In Weapon Release authorization task ( $d = -.62$ ), task demand had a stronger effect on dependence on automation than in Image Analysis task ( $d = -.03$ ). Specifically, in Weapon Release authorization task, participants were less dependent on automation in high task demand conditions than in low task demand conditions.



**Fig. 5.** Task performance (dependence on automation) in the Image Analysis and the Weapon Release authorization tasks for different task demand conditions. Error bars represent standard errors.

### *Neglect*

Regarding neglect, there was significantly more item neglect instances in the Weapon Release authorization task ( $M = 8.9$ ) than in the Image Analysis task ( $M = 3.4$ ),  $F(1, 91) = 94.08$ ,  $p < .01$ ,  $\eta_p^2 = .51$ . The main effects for task demand and LOA were also significant for neglect (see Fig. 6). First, neglect was higher in high task demand groups ( $M = 8.4$ ) than in low task demand groups ( $M = 3.9$ ),  $F(1, 91) = 19.18$ ,  $p < .01$ ,  $\eta_p^2 = .17$ . Second, neglect was higher in high LOA conditions ( $M = 7.1$ ) than in low LOA conditions ( $M = 5.1$ ),  $F(1, 91) = 4.20$ ,  $p < .05$ ,  $\eta_p^2 = .04$ . Although significant, the effects of LOA manipulations on neglect were small in both Weapon Release authorization ( $d = .31$ ) and Image Analysis ( $d = .35$ ) tasks. In addition, the interaction between task type and task demand was significant,  $F(1, 91) = 9.68$ ,  $p < .01$ ,  $\eta_p^2 = .10$ . The effect of task demand had a stronger impact on Weapon Release authorization task ( $d = .92$ ) than on Image Analysis task ( $d = .59$ ). Participants in the high task demand conditions neglected the most number of items in the Weapon Release authorization task.



**Fig. 6.** Task performance (neglect) in the Image Analysis and the Weapon Release authorization tasks for different task demand conditions. Error bars represent standard errors.

## 4 Discussion

A main object of this study was apply a multivariate assessment strategy [20] to profile the workload, stress and performance changes associated with high task demands during simulated UAS operation. Supporting this objective, NASA-TLX, DSSQ, and performance-based measures were implemented in this study. Consistent with the SCM [12], the task demand manipulations elicited higher subjective distress and workload, as well as performance deficits. However, there was considerable variation in effect size across different indicators of overload, supporting the multivariate approach to assessing task demands impact. Other aspects of response differed qualitatively from expectation: specifically, LOAs did not affect operator workload, but affected dependence on automation.

The remainder of this section will evaluate task demand impacts on the various subjective and objective outcome measures. Most, though not all, task demand effects were consistent with prediction from the SCM [12], but variation in effect sizes indicates the limitations of relying on the capacity model alone in mitigating overload effects in the operational setting. Findings also indicated the role of participant strategy, for example, in choosing to neglect task items, and reducing automation-dependence under overload. Maladaptive strategy choices of these kinds may be addressed in training for management of overload.

### 4.1 Subjective Measures

Workload was assessed subjectively via NASA-TLX. Due to the convenience for administration and analysis, subjective measures are the most commonly applied method to assess workload. NASA-TLX, as one of the most popular subjective measures, assesses multiple aspects of workload. It is sensitive to overall task load and diagnostic of the nature of workload from six separate sources [20]. In this study, the level of task demand was successfully manipulated to simulate the task demand variation in UAS operations by configuring the frequency of secondary task events in the ALOA simulation. Higher subjective workload was reported by the high task demand group. According to Cohen's [31] guidelines, the 0.70 effect size for overall ratings was close to the standard for a large effect (0.80). But effect sizes varied in the individual ratings. For example, the effect sizes in physical demand, temporal demand, and frustration were medium-magnitude, whereas the effect size in mental demand was even smaller. Based on the individual ratings, physical demand was the primary source of subjective workload, presumably because frequent mouse movements were required. Therefore, designing interfaces to reduce manual demands may contribute to mitigating overall workload under high demand in UAS missions. Improvements to the interface might also reduce frustration. Time pressure is harder to mitigate because it reflects external contingencies beyond the operator's control, although future systems with greater automation of decision-making might relieve the temporal burden.

In addition, the LOA manipulations did not affect perceived workload as expected, on the basis that higher LOAs should free up attentional capacity. A possible explanation is that two intermediate levels, management-by-consent and management-by-exception, were selected from the LOA model [8] in the study. These two levels may

be too close to make a profound difference in the effect of LOA on workload. Alternatively, at the higher LOA, the operator may have reallocated attention to additional activities, such as secondary tasks, so that workload remained constant.

Subjective states of task stress were assessed via the DSSQ. This scale is widely used for profiling state change induced by task and environmental stressors in basic and applied performance tasks [24]. Using the DSSQ makes it possible to compare state change profiles across studies in the same domain. Consistent with previous studies of simulated unmanned system operations, high task demand produced greater distress, i.e., negative emotions and loss of confidence in performance [21, 22, 30]. Elevated task engagement was also reported in the higher task demand conditions, implying that the greater workload may have helped operators maintain motivation and alertness.

However, the changes in subjective state were only of small effect size, indicate that individuals can cope effectively despite high workload. Distress is normally associated with higher workload [25], but the change in distress produced by high task demands here was associated with an effect size of only .30. By contrast, Abich et al. [22] found a dual-tasking effect size of .97 in a UGV simulation study. In Abich et al.'s [22] study, participants performed surveillance tasks only, whereas ALOA requires participants to multi-task a more diverse collection of subtasks, which may have provided a more interesting and challenging assignment. Support from automation in ALOA, not present in Abich et al.'s [22] tasks, may also have limited distress. Therefore, in MAC stress intrinsic to task demands may be a minor concern operationally, although no attempt was made to simulate additional stress factors that may be present in real operations [11].

Worry was reduced relative to baseline in both task demand conditions. Typically, demanding tasks can induce decreases in worry, as attention is refocused from internal concerns to external demands [25]. The present result was also consistent with the trend of greater declines in worry in high event rate vigilance tasks [32]. By contrast, low workload, monotonous UAS tasks may lead to mind-wandering, which may, in turn, contribute to increases in worry in long-duration missions [28].

## 4.2 Performance-Based Measures

Three performance metrics, including accuracy, neglect, and dependence on automation, were developed for the primary surveillance tasks. Measuring task performance, such as accuracy and neglect, can provide an indication of workload and stress, especially when the operator is overloaded and performance is impaired. Performance can be impaired when demands exceed the operator's limited attentional resources [12]. Dependence on automation is an indicator of operator trust in automation. It helps to profile how operators utilize automation in coping with overloaded situations and how trust is impacted by overload.

Generally, task demand manipulations impacted performance as expected, except some findings regarding dependence on automation. Overall, larger effect sizes were found for the Weapon Release task across all three performance metrics, compared with Image Analysis. From the perspective of the SCM [12], the Weapon Release task was more demanding and required more attentional resources. Therefore, it was more vulnerable to high task demands. For both tasks, effect sizes tended to be larger for

neglect than for accuracy, implying that participants may have tried to compensate for demands by reducing the number of images they had to process, especially for Weapon Release.

In addition, dependence on automation in the Image Analysis task was consistent across task demand conditions, while significantly less dependence on automation in the Weapon Release task was observed in the high task demand condition. The moderate-magnitude effect of task demand in the Weapon Release task was unexpected. An attentional capacity perspective [29] would suggest that automation dependence should increase under higher task demands as the operator's attention is increasingly taxed. High task demands may not only deplete the limited cognitive resource, but impair operators' trust in automation. Such effects may also reflect operators' tendency to adopt task-focused coping as a strategy, relying on personal agency rather than automation for dealing with the overloaded situation. In fact, this strategy seems counterproductive, and training for operators might focus on the need to trust automation in overload situations.

Even though no effect of LOA on subjective workload and stress states was found, differences in performance-based measures were observed between the two LOA configurations. Greater dependence on automation and more neglect were observed in higher LOA conditions (management-by-exception). Higher LOA may lead to a loss of situation awareness associated with vigilance decrement and complacency issues [33] and may, in turn, result in the observed greater dependence on automation and more neglect. No significant difference in task accuracy was found between LOA conditions. This finding may suggest that considering the automation is relatively reliable, LOAs only have a subtle effect on the overall accuracy even though higher LOAs encourage operators to rely on the automation more. Also, the two LOAs were at intermediate levels close to each other. Future study may test the trend at other LOAs.

### 4.3 Limitations and Future Work

One limitation is the use of novice participants rather than trained operators. Applying the SCM to expertise suggests that increased skill will reduce capacity demands, and hence the vulnerability of the operator to both cognitive overload and stress [12]. However, even skilled operators remain vulnerable to overload in some circumstances, and multivariate assessment of response remains a useful methodology for guiding mitigation strategies [24]. Nevertheless, it would be desirable to further examine the roles of practice and expertise in moderating overload impacts.

Also, subjective measures are highly applicable to assessing an operator's workload when interacting with modern technologies that aid judgment and decision making [34], such as in UAS domains, and are very useful to assess stress states for testing predictions from theory [24]. But subjective measures alone are inadequate to effectively characterize workload because they can become insensitive to changes in task demand [35]. In addition, subjective measures are not applicable for diagnostic monitoring during a mission. Therefore psychophysiology assessments are also necessary for better understanding, predicting, and improving human performance in complex human-machine systems, such as UAS operations. Psychophysiology measures, such as electrocardiogram (ECG), electroencephalogram (EEG), transcranial Doppler

sonography (TCD), functional near-infrared spectroscopy (fNIR), and eye tracking, are particularly useful in assessing workload and stress state and are applicable for operational diagnostic monitoring [36].

**Acknowledgement.** This research was sponsored by AFOSR A9550-13-1-0016 and 13RH05COR. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of AFOSR or the US Government.

## References

1. Calhoun, G.L., Ruff, H.A., Draper, M.H., Wright, E.J.: Automation-level transference effects in simulated multiple unmanned aerial vehicle control. *J. Cogn. Eng. Decis. Making* **5**(1), 55–82 (2011)
2. Guznov, S., Matthews, G., Funke, G., Dukes, A.: Use of the RoboFlag synthetic task environment to investigate workload and stress responses in UAV operation. *Behav. Res. Methods* **43**(3), 771–780 (2011)
3. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. *Hum. Factors* **46**(1), 50–80 (2004)
4. Hancock, P.A., Warm, J.S.: A dynamic model of stress and sustained attention. *Hum. Factors* **31**(5), 519–537 (1989)
5. Dixon, S.R., Wickens, C.D., Chang, D.: Mission control of multiple unmanned aerial vehicles: a workload analysis. *Human Factors* **47**(3), 479–487 (2005)
6. Hart, S.G., Wickens, C.D.: Cognitive workload in NASA human integration design handbook (NASA/SP-2010-3407). NASA, Washington, DC (2010)
7. Lee, J.D.: Affect, attention, and automation. In: Kramer, A., Wiegmann, D., Kirlik, A. (eds.) *Attention: From Theory to Practice*, pp. 73–89. Oxford University Press, New York (2006)
8. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A model for types and levels of human interaction with automation. *IEEE Trans. Syst. Cybern. Part A Syst. Hum.* **30**(3), 286–297 (2000)
9. Lewis, M.: Human interaction with multiple remote robots. *Rev. Hum. Factors Ergon.* **9**(1), 121–174 (2013)
10. Matthews, G., Campbell, S.E., Falconer, S., Joyner, L.A., Huggins, J., Gilliland, K., Grier, R., Warm, J.: Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion* **2**(4), 315–340 (2002)
11. Ouma, J.A., Chappelle, W.L., Salinas, A.: Facets of occupational burnout among U.S. Air Force active duty and national guard/reserve MQ-1 Predator and MQ-9 Reaper operators (AFRL-SA-WP-TR-2011-0003). School of Aerospace Medicine Wright Patterson AFB OH (2011)
12. Matthews, G., Wohleber, R.W., Lin, J.: Stress, skilled performance, and expertise: Overload and beyond. In: Ward, P., Schraagen, J.M., Gore, J., Roth, E. (eds.) *The Oxford Handbook of Expertise*. Oxford University Press, New York (in press)
13. Wohleber, Ryan W., Matthews, G., Funke, Gregory J., Lin, J.: Considerations in physiological metric selection for online detection of operator state: a case study. In: Schmorrow, Dylan D.D., Fidopiastis, Cali M.M. (eds.) *AC 2016. LNCS (LNAI)*, vol. 9743, pp. 428–439. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-39955-3\\_40](https://doi.org/10.1007/978-3-319-39955-3_40)
14. Warm, J.S., Parasuraman, R., Matthews, G.: Vigilance requires hard mental work and is stressful. *Hum. Factors* **50**(3), 433–441 (2008)

15. Warm, J.S., Matthews, G., Finomore, V.S.: Workload and stress in sustained attention. In: Hancock, P.A., Szalma, J.L. (eds.) *Performance Under Stress*, pp. 115–141. Ashgate Publishing, Aldershot (2008)
16. Raby, M., Wickens, C.D.: Strategic workload management and decision biases in aviation. *Int. J. Aviat. Psychol.* **4**(3), 211–240 (1994)
17. Parasuraman, R., Riley, V.: Humans and automation: use, misuse, disuse, abuse. *Hum. Factors* **39**(2), 230–253 (1997)
18. Dixon, S.R., Wickens, C.D., McCarley, J.S.: On the independence of compliance and reliance: Are automation false alarms worse than misses? *Hum. Factors* **49**(4), 564–572 (2007)
19. Barg-Walkow, L.H., Rogers, W.A.: The effect of incorrect reliability information on expectations, perceptions, and use of automation. *Hum. Factors* **58**(2), 242–260 (2016)
20. Matthews, G., Reinerman-Jones, L.E.: Workload assessment: How to diagnose workload issues and enhance performance. In: *Human Factors and Ergonomics Society*, Santa Monica, CA (2017)
21. Panganiban, A.R., Matthews, G.: Executive functioning protects against stress in UAV simulation. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **58**(1), 994–998 (2014)
22. Abich IV, J., Reinerman-Jones, L.E., Matthews, G.: Impact of three workload factors on simulated unmanned system intelligence, surveillance, and reconnaissance operations. *Ergonomics* **60**(6), 791–809 (2017)
23. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) *Advances in Psychology*, North-Holland, Amsterdam, The Netherlands, vol. 52, pp. 139–183 (1988)
24. Matthews, G.: Multidimensional profiling of task stress states for human factors: a brief review. *Hum. Factors* **58**(6), 801–813 (2016)
25. Matthews, G., Szalma, J.L., Panganiban, A.R., Neubauer, C., Warm, J.S.: Profiling task stress with the Dundee Stress State Questionnaire. In: Cavalcanti, L., Azevedo, S. (eds.) *Psychology of Stress: New Research*, pp. 49–90. Nova, Hauppauge (2013)
26. Matthews, G., Warm, J.S., Reinerman-Jones, L.E., Langheim, L.K., Washburn, D.A., Tripp, L.: Task engagement, cerebral blood flow velocity, and diagnostic monitoring for sustained attention. *J. Exp. Psychol. Appl.* **16**(2), 187–203 (2010)
27. Guznov, S., Matthews, G., Warm, J.S., Pfahler, M.: Training techniques for visual search in complex task environments. *Hum. Factors* **59**(7), 1139–1152 (2017)
28. Cummings, M.L., Mastracchio, C., Thornburg, K.M., Mkrtchyan, A.: Boredom and distraction in multiple unmanned vehicle supervisory control. *Interact. Comput.* **25**(1), 34–47 (2013)
29. Parasuraman, R., Manzey, D.H.: Complacency and bias in human use of automation: an attentional integration. *Hum. Factors* **52**(3), 381–410 (2010)
30. Wohleber, R.W., Matthews, G., Reinerman-Jones, L.E., Panganiban, A.R., Scribner, D.: Individual differences in resilience and affective response during simulated UAV operations. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **59**(1), 751–755 (2015)
31. Cohen, J.: A power primer. *Psychol. Bull.* **112**(1), 155–159 (1988)
32. Shaw, T.H., Matthews, G., Warm, J.S., Finomore, V.S., Silverman, L., Costa Jr., P.T.: Individual differences in vigilance: personality, ability and states of stress. *J. Res. Pers.* **44**(3), 297–308 (2010)
33. Endsley, M.R., Kiris, E.O.: The out-of-the-loop performance problem and level of control in automation. *Hum. Factors* **37**(2), 381–394 (1995)
34. Cain, B.: A review of the mental workload literature. Technical report, Defense Research and Development Toronto, Canada (2007)



35. Wierwille, W.W.: Important remaining issues in mental workload estimation. In: Hancock, P.A., Meshkati, N. (eds.) *Human Mental Workload*, pp. 315–327. North-Holland, Amsterdam (1988)
36. Matthews, G., Reinerman-Jones, L.E., Barber, D.J., Abich IV, J.: The psychometrics of mental workload: multiple measures are sensitive but divergent. *Hum. Factors* **57**(1), 125–143 (2015)