2007

Parameters Affecting Mental Workload and the Number of Simulated UCAVS that can be Effectively Supervised

Bryan A. Calkin
Wright State University

Follow this and additional works at: http://corescholar.libraries.wright.edu/etd_all

Part of the Industrial and Organizational Psychology Commons

Repository Citation
Calkin, Bryan A., "Parameters Affecting Mental Workload and the Number of Simulated UCAVS that can be Effectively Supervised" (2007). Browse all Theses and Dissertations. Paper 104.

This Thesis is brought to you for free and open access by the Theses and Dissertations at CORE Scholar. It has been accepted for inclusion in Browse all Theses and Dissertations by an authorized administrator of CORE Scholar. For more information, please contact corescholar@www.libraries.wright.edu.
PARAMETERS AFFECTING MENTAL WORKLOAD AND THE NUMBER OF SIMULATED UCAVS THAT CAN BE EFFECTIVELY SUPERVISED

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science

By

BRYAN A. CALKIN
B.S., Troy University, 2002

2007
Wright State University
I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Bryan A. Calkin ENTITLED Parameters Affecting Mental Workload and the Number of Simulated UCAVs that can be Effectively Supervised BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science.

Herbert A. Colle, Ph.D.
Thesis Director

Kevin B. Bennett, Ph.D.
Graduate Program Director

John M. Flach, Ph.D.
Department Chair

Committee on Final Examination

Herbert A. Colle, Ph.D.

Valerie L. Shalin, Ph.D.

Christopher A. Russell, Ph.D.

Joseph F. Thomas, Jr., Ph.D.
Dean, School of Graduate Studies
ABSTRACT

Calkin, Bryan A. M.S., Department of Psychology, Wright State University, 2007. Parameters Affecting Mental Workload and the Number of Simulated UCAVs that can be Effectively Supervised

As Unmanned Combat Aerial Vehicles (UCAVs) become integrated into the U.S. military’s arsenal, the number of vehicles that an operator can successfully supervise will play an important role in the effectiveness of future missions. The present study investigated performance and mental workload when an operator supervises multiple UCAVs. This study focused on the parameters that affect the operator’s performance during a simulated UCAV suppression of enemy air defenses (SEAD) mission, which is expected to be the primary function of the UCAV. All three factors which were manipulated, including the number of vehicles to be supervised, vehicle airspeed, and difficulty level of attacks (targets engaged by either a single vehicle or multiple vehicles), affected both performance and Subjective Workload Assessment Technique (SWAT), and NASA Task Load Index (TLX), subjective mental workload measures. Accomplishment score analyses were used to estimate performance redlines, based on the Accomplishment Score Model of Average Mental Workload (Colle & Reid, 1997, 2005). A performance mental workload redline was defined as the point at which accomplishment scores no longer increased. Performance redlines were estimated using piecewise linear functions of accomplishment scores. Redlines indicated that for simple scenarios operators could effectively control about 12 UCAVs flying at 900 knots or 8 UCAVs flying at 1500 knots. For complex scenarios, operators could effectively control 8 UCAVs flying at 900 knots.

Subjective mental workload redlines also were estimated for both the SWAT and TLX subjective mental workload measures based on the performance redlines. Consistent with
previous research, the estimated SWAT redline was in the range of 40 ± 10. Initial estimates of a redline also were obtained for the TLX.
# TABLE OF CONTENTS

I. Introduction .................................................................................................................. 1

II. Mental workload ........................................................................................................... 5
   a. Single resource mental workload theories ............................................................... 5
   b. Types of measures ...................................................................................................... 8
      i. Primary task performance ..................................................................................... 9
      ii. Subjective ............................................................................................................. 10
      iii. Mental Workload Redline .......................................................... 13
      iv. Secondary task performance .............................................................................. 13

III. Goals of the Study ...................................................................................................... 13

IV. Method ........................................................................................................................ 17
   a. Participants ................................................................................................................ 17
   b. Facilities, software, and apparatus .......................................................................... 17
   c. Scenario .................................................................................................................... 18
   d. Experimental design ................................................................................................. 27
   e. Procedure .................................................................................................................. 30
   f. Statistical analyses .................................................................................................... 34

V. Results .......................................................................................................................... 37
   a. Performance ............................................................................................................... 37
      i. Percent Kills .......................................................................................................... 37
      ii. Percent DMPIs Correctly Placed ......................................................................... 39
   b. Subjective ratings ..................................................................................................... 43
      i. SWAT ...................................................................................................................... 43
      ii. NASA TLX ............................................................................................................ 45
   c. Estimation of Maximum Performance: Performance Redlines ............................. 47
      i. High Attack Difficulty ........................................................................................ 49
      ii. Low Attack Difficulty ......................................................................................... 53
   d. Estimation of Subjective Mental Workload Redlines ........................................... 55
      i. High Difficulty – SWAT .................................................................................... 55
      ii. Low Difficulty – SWAT ..................................................................................... 56
      iii. High Difficulty – TLX-M4 .............................................................. 58
iv. Low Difficulty – TLX-M4

VI. Discussion
   a. Summary of Research Goals
   b. Limitations and Future Research
   c. Potential Applications

VII. References

Appendix A: NASA TLX Dimensions
Appendix B: SWAT Dimensions
Appendix C: TLX-M4 and TLX - All Factors Rating Scales
Appendix D: Training Procedure
Appendix E: Numeric SWAT scale estimated from the card sort and conjoint analysis
Appendix F: Summary of Distracters Destroyed
Appendix G: ANOVA on Percent Kills for High and Low Difficulty
Appendix H: ANOVA on Percent Kills for High Difficulty
Appendix I: ANOVA on Percent Kills for Low Difficulty
Appendix J: ANOVA on Percent DMPIs for High and Low Difficulty
Appendix K: ANOVA on Percent DMPIs for High Difficulty
Appendix L: ANOVA on Percent DMPIs for Low Difficulty
Appendix M: ANOVA on SWAT Rating for High and Low Difficulty
Appendix N: ANOVA on TLX Rating All Factors for High and Low Difficulty
Appendix O: ANOVA on TLX-M4 Rating for High and Low Difficulty
Appendix P: ANOVA on Kills Accomplishment Score for High Difficulty
Appendix Q: Piecewise Regression Best-Fitting Estimations for High Difficulty Kills and DMPIs: Splines & knots
Appendix R: Piecewise Regression Best-Fitting Estimations for Low Difficulty Kills and DMPIs: Splines & knots
Appendix S: ANOVA on DMPIs Accomplishment Score for High Difficulty
Appendix T: ANOVA on Kills Accomplishment Score for Low Difficulty
Appendix U: ANOVA on DMPIs Accomplishment Score for Low Difficulty
Appendix V: ANOVA on SWAT Redline Rating Accomplishment Score for High Difficulty
Appendix W: ANOVA on SWAT Redline Rating Accomplishment Score for Low Difficulty
<table>
<thead>
<tr>
<th>Appendix</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>ANOVA on TLX-M4 Redline Rating Accomplishment Score for High Difficulty</td>
<td>98</td>
</tr>
<tr>
<td>Y</td>
<td>ANOVA on TLX-M4 Redline Rating Accomplishment Score for Low Difficulty</td>
<td>99</td>
</tr>
<tr>
<td>Z</td>
<td>ANOVA on TLX Rating All Factors Accomplishment Score for High Difficulty</td>
<td>100</td>
</tr>
<tr>
<td>AA</td>
<td>ANOVA on TLX Rating All Factors Accomplishment Score for Low Difficulty</td>
<td>101</td>
</tr>
<tr>
<td>AB</td>
<td>Summary of SWAT Axiom Tests</td>
<td>102</td>
</tr>
<tr>
<td>AC</td>
<td>Summary of Mean Raw NASA TLX Ratings by Subscale</td>
<td>103</td>
</tr>
<tr>
<td>AD</td>
<td>Summary of Percent Kills Data</td>
<td>104</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Mental workload regions as described by O’Donnell and Eggemeier</td>
<td>8</td>
</tr>
<tr>
<td>2.</td>
<td>Two strike packages shown on TSD</td>
<td>19</td>
</tr>
<tr>
<td>3.</td>
<td>UCAVs engaged with target icon</td>
<td>20</td>
</tr>
<tr>
<td>4.</td>
<td>Examples of valid targets</td>
<td>21</td>
</tr>
<tr>
<td>5.</td>
<td>Examples of distracters</td>
<td>22</td>
</tr>
<tr>
<td>6.</td>
<td>A typical low difficulty SAR image</td>
<td>24</td>
</tr>
<tr>
<td>7.</td>
<td>A typical high difficulty SAR image</td>
<td>25</td>
</tr>
<tr>
<td>8.</td>
<td>Percent Kills for High Attack Difficulty</td>
<td>38</td>
</tr>
<tr>
<td>9.</td>
<td>Percent Kills for Low Attack Difficulty</td>
<td>38</td>
</tr>
<tr>
<td>10.</td>
<td>Percentage of correctly-placed DMPIs for High Attack Difficulty Scenarios</td>
<td>42</td>
</tr>
<tr>
<td>11.</td>
<td>Percentage of correctly-placed DMPIs for Low Attack Difficulty Scenarios</td>
<td>42</td>
</tr>
<tr>
<td>12.</td>
<td>SWAT Ratings for High Attack Difficulty Scenarios</td>
<td>44</td>
</tr>
<tr>
<td>13.</td>
<td>SWAT Ratings for Low Attack Difficulty Scenarios</td>
<td>44</td>
</tr>
<tr>
<td>14.</td>
<td>TLX-M4 Ratings for High Attack Difficulty Scenarios</td>
<td>46</td>
</tr>
<tr>
<td>15.</td>
<td>TLX-M4 Ratings for Low Attack Difficulty Scenarios</td>
<td>46</td>
</tr>
<tr>
<td>16.</td>
<td>Accomplishment Kill score for High Attack Difficulty Scenarios</td>
<td>52</td>
</tr>
<tr>
<td>17.</td>
<td>Accomplishment DMPI score for High Attack Difficulty Scenarios</td>
<td>53</td>
</tr>
<tr>
<td>18.</td>
<td>Accomplishment Kill score for Low Attack Difficulty Scenarios</td>
<td>54</td>
</tr>
<tr>
<td>19.</td>
<td>Accomplishment DMPI scores for Low Attack Difficulty Scenarios</td>
<td>55</td>
</tr>
<tr>
<td>20.</td>
<td>SWAT redlines as indicated from high difficulty Performance Redline</td>
<td>56</td>
</tr>
<tr>
<td>21.</td>
<td>SWAT redlines as indicated from low difficulty Performance Redline</td>
<td>57</td>
</tr>
<tr>
<td>22.</td>
<td>TLX-M4 redlines as indicated from high difficulty Performance Redline</td>
<td>58</td>
</tr>
<tr>
<td>23.</td>
<td>TLX-M4 redlines as indicated from low difficulty Performance Redline</td>
<td>60</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Mental workload regions as described by O’Donnell and Eggemeier</td>
</tr>
<tr>
<td>2.</td>
<td>Randomized order of operator trials</td>
</tr>
<tr>
<td>3.</td>
<td>Description of how DMPIs and Kills were scored</td>
</tr>
<tr>
<td>4.</td>
<td>Sequence of Scenarios and Maximum Possible Numbers of Kills and DMPIs used for the Accomplishment Score Analyses</td>
</tr>
</tbody>
</table>
Acknowledgements

First and foremost, I would like to thank my wife Valerie for supporting me during this process with her unending love, patience, and confidence in me. I would also like to thank my committee members, Dr. Herb Colle, Dr. Chris Russell, and Dr. Valerie Shalin for their guidance in the completion of this research. Special thanks also go to Dr. Glenn Wilson and Dr. Grant McMillan for allowing me to use the resources in the Flight Psychophysiology Laboratory to complete this study.
I. Introduction

As Unmanned Combat Aerial Vehicles (UCAVs) become integrated into the U.S. military, the number of vehicles that an operator can successfully supervise will play an important role in the effectiveness of future combat missions. In current unmanned aerial vehicle (UAV) operations, two basic roles exist: a pilot/supervisor and a sensor operator. At present, these responsibilities are divided among two or more people per aircraft. Existing large-scale UAV platforms such as the Predator and Global Hawk are used primarily for intelligence, surveillance, and reconnaissance (ISR).

The present study focused on the parameters that affect operator’s performance during a simulated UCAV suppression of enemy air defenses (SEAD) mission, which is envisioned as the primary function of the UCAV. Other proposed UCAV functions include electronic attack, penetrating surveillance, and related missions that require timely strike capabilities (Francis, 2005). While the ISR function also requires UAV operators to use supervisory control in a multi-task environment, the combat role brings with it an added burden of responsibility to ensure that decisions of targeting and release of weapons are as accurate as possible to reduce unnecessary loss of life and collateral damage.

In an effort to address these challenges, as well as reducing the number of operators that are required for controlling these vehicles, the U.S. military is seeking to identify the sources of mental workload that influence the effectiveness of the operators. The factors that contribute to determining how many remotely operated vehicles one operator can optimally manage are still being explored (Wickens, Dixon, & Chang, 2003; Dolgin, Kay, Langelier, Wasel, & Hoffmann, 2002; Dixon & Wickens, 2003; Nelson, Lefebvre, & Andre, 2004; Cummings & Guerlain, 2004; Wilson & Russell, 2004). Previous research in the UAV domain has suggested that performance
begins to degrade when a single operator manages a small number of vehicles, approximately 3-5 (Nelson, Lefebvre, & Andre, 2004; Ruff, Calhoun, Draper, Fontejon, & Guilfoos, 2004; Wilson & Russell, 2004).

While several studies have explored the effective operator-to-vehicle ratio (see McCarley & Wickens, 2005), the Air Force Research Laboratory, Flight Psychophysics Laboratory's (FPL) preliminary data, including those collected during a collaboration between the Boeing Company and AFRL under the Defense Advanced Research Projects Agency’s (DARPA) Augmented Cognition (AugCog) program, suggest that managing up to twelve vehicles is achievable (Barker, 2004, 2005). During these concept validation experiments (CVEs), the experimental participants were able to achieve these results by using the simulator’s adaptive automation, which was driven by psychophysiological sensor inputs.

One of the FPL laboratory’s goals during the AugCog program was to study the ability of physiological sensors to capture the mental workload-related bio-data, and then accurately reduce, synthesize, and integrate the biological signals into a real-time estimation of operator functional state (OFS). This real-time assessment of operator functional state was then used to initiate and withdraw the adaptive automation in a closed-loop environment (Wilson, 2003; Wilson & Russell, 2004). Previous research in the AugCog program has utilized dual-task methods in which the operators performed both SEAD targeting functions and vehicle health tasks, which occurred intermittently throughout each mission scenario. The present study focused on targeting only, because: (a) whether one operator handles both tasks or whether a separate operator handles all health and maintenance issues is a system design and implementation concern, and (b) there is a variety of additional tasks that operators might be required to perform in the final operationally-fielded system. The workload demands of these
additional tasks and the manner in which they are distributed are presently unspecified. Additionally, knowing the primary task performance range allows dual-task implications to be evaluated more systematically. Tsang and Shaner (1998) described the necessity of establishing a single task baseline, especially for dual-task studies which investigate the effects of expertise and aging on time-sharing performance.

Primary task performance of UCAV operators will be notably affected by several key factors. The interface that enables the interaction between the system and the operator will serve as a crucial element in these missions. Due to the remote nature of the UCAV, displays will serve as the primary or sole source of information concerning the vehicles and their activities (Flach, Eggleston, Kuperman, & Dominguez, 1998). This physical separation of operator and vehicle will also present numerous challenges to effective operator performance, through loss of sensory cues, communication and control delays, and reduced visual and sensory awareness (McCarley & Wickens, 2005).

Additionally, the operators’ mental workload demands during these dynamic missions will fluctuate, and in turn, influence their effectiveness. Long periods of low mental workload, during ingress and egress (under auto-pilot or “waypoint-to-waypoint” supervisory control), will be interrupted by the high task demands encountered in the area of responsibility (AOR), as well as the takeoff and landing processes. These AOR tasks will likely include analyzing imagery, placing designated mean points of impact (DMPIs), authorizing munitions for release and communicating with other manned and unmanned assets on the battlefield.

The present simulated UCAV multi-task environment study manipulated three independent variables: (a) number of vehicles to be supervised, (b) vehicle airspeed, and (c) difficulty level of attacks (targets engaged by either a single vehicle or multiple vehicles). The
effects of both primary task performance and mental workload, assessed by subjective measures, were examined. A primary goal of the present study was to estimate the number of UCAVs one operator can manage as a function of airspeed and task difficulty. Given that the issues involving adaptive automation for remotely operated vehicles have yet to be fully resolved, another goal of the present study was to develop performance guidelines for the systems that are initially fielded without this automation. An additional goal was to provide baseline performance data in this new task simulation environment that can contribute to follow-on studies that will examine adaptive automation.
II. Mental Workload

a. **Single Resource Theories**

Several theories presume that mental workload is fundamentally a single dimension or resource (Hancock & Caird, 1993; Hendy, Liao, & Milgram, 1997; Colle & Reid, 1997). These theories assume that selective interactions between different types of tasks may exist because of what Kahneman (1973) called structural interference. Most of these theories are process-oriented, describing the particular characteristics with assumptions about the underlying, unobservable cognitive processes. One theory, the accomplishment model of average mental workload, as described by Colle and Reid (1997), takes a different approach. It is a fundamental measurement theory as outlined by Krantz, Luce, Suppes, and Tversky (1971). Colle and Reid (1997) stated that, conceptually, the accomplishment scores represent an index of the amount of work accomplished in a task.

This extensive measurement-based theory contains three elements: (a) mental workload entities, (b) a comparison-equivalence relationship, and (c) an empirical concatenation operation. First, mental workload entities are analogous to the objects one might use as entities in mass measurements, comparing one object with another, for example, on a pan balance. Colle and Reid (1997) argued that to measure mental workload, entities should be defined by both the cognitive task and the performance accomplished on it during a fixed time interval. They also argued that mental workload entities should be empirically defined and identifiable. To measure mental workload performance variables that capture the amount of mental work completed during the fixed time interval, accomplishment scores. Second, a comparison-equivalence relationship must be defined. It could consist of providing a mental workload entity stimulus and comparing it to a given standard. An analogous example from psychophysics might be to
present a participant with a visible wavelength stimulus at a fixed intensity and asking participants to compare, or equate, the stimulus to a standard wavelength and intensity. Lastly, an empirical concatenation operation is needed. A concatenation operation allows two entities to be combined so that their sum can be evaluated by the comparison relationship. Colle and Reid (1997) described how mental workload needed to be estimated from accomplishment scores, the average amount of work completed or successfully performed, in a fixed time period.

Concatenation of mental workload entities is defined over this same time interval. This interval could be considered like a mental workload “time bin.” The bin has time on the x-axis and instantaneous mental workload on the y-axis. Thus, the volume of the bin is the total mental work that can be accomplished during the interval. When a task is accomplished during the interval it has a profile of instantaneous mental workload levels. So, a set of cognitive task can be performed as long as their total work does not exceed the maximum total work for the interval. Concatenation of mental workload entities populates the bin with successfully accomplished tasks. It is assumed that there is a maximum amount of mental work that the bin can hold.

As a consequence of the assumptions made by the accomplishment model, Colle and Reid (1997) indicated that the information processing stages of tasks may include serial processes, parallel processes, or both. The accomplishment model assumes only one dimension of mental workload. This model only applies to average mental workload and to certain types of performance measures and dependent variables that summarize performance in terms of the overall amount of cognitive work accomplished. By specifying a broader, fixed-time interval and measuring the number of successfully completed tasks, the processing stages and resources are allowed to be shared, redistributed and re-prioritized. According to the accomplishment
model of assessing mental workload, secondary task measures are only appropriate at the lower mental workload levels prior to the performance redline (Colle & Reid, 2005).

O’Donnell and Eggemeier (1986) described three distinct regions of mental workload in their hypothetical relationship between mental workload and performance. In Region A, low to moderate levels of mental workload leave the operator with sufficient resources or processing capacity to compensate for increases in task demand. Performance is good throughout the range. In Region B, the higher mental workload levels outpace the cognitive capabilities of the operator. Degradations in performance begin to occur, and a monotonic relationship exists between mental workload and performance. In Region C, very high levels of mental workload result in poor performance, as a consequence of the operator’s shortage of cognitive capacity. The transition from Region A to Region B was considered to be a redline point by Colle and Reid (2005). They empirically defined a performance redline as an estimation of this transition point. Figure 1 shows the three regions and a hypothetical performance curve, using a typical measure such as percent correct, as operator workload increases. The arrow seen in Figure 1 indicates the performance redline. The present study focused on investigating the parameters that determine when UCAV operator performance is above or below the performance redline.
b. Types of Measures

Gopher and Donchin (1986) referred to mental workload as a hypothetical construct that is intended to capture limitations on the information processing capabilities of the operator, while engaged in some task. They also described how people are able to apply varying levels of mental and physical effort to maintain stable performance during a task as it fluctuates in difficulty. This can complicate the mental workload measurement process, especially in dynamic multi-task environments.
A single, universally-accepted definition of mental workload does not exist. Many have been proposed, but the current set of definitions has not stood the test of general acceptance or quantitative validation (Xie and Salvendy, 2000a). Most definitions involve the interaction of two fundamental components: a person and a task. There are more specific elements that combine to affect this interaction, such as mental effort, cognitive resources, cognitive ability, and time. Xie and Salvendy (2000b) suggested that while mental workload cannot be directly detected, it can be measured through other means that are thought to correlate highly with it, such as performance indicators, subjective ratings and psychophysiological measures. Tsang and Wilson (1997) suggested that because workload is a multifaceted construct, using multiple workload measures should provide a more well-rounded assessment of the task demands. Some commonly-used types of mental workload measures are discussed below.

Primary Task Performance. O’Donnell and Eggemeier (1986) stated that primary task measures are frequently used to assess mental workload. Their basic assertion was that as additional mental processing resources were applied, some discernible effect on the quality of operator performance would become evident. By measuring these effects, an index of mental workload could be obtained. Some limitations exist in this method. In particular, the area of low workload (i.e., Region A) may not show performance decrements, even though mental effort may increase as task demands fluctuate.

One major use of primary task performance measures has been to distinguish the separation between a non-overload and an overload condition. This transition from Region A to Region B is shown in Figure 1. Because the specific demands of the UCAV operator’s primary task demands (i.e., processing imagery for time-critical release of weapons) have not yet been well established, basic examinations of this process are still required. Identifying the point at
which performance begins to suffer is critical to the real-world effectiveness of the UCAV mission. Given that task demands while supervising multiple vehicles are expected to exceed the operator’s performance redline, performance on the primary task of processing target images served as an indicator of mental workload.

Subjective. Most subjective measures of mental workload were developed independently of existing mental workload theories (Wierwille & Eggemeier, 1993). However, they are the most commonly used tools for practical mental workload evaluation. Using subjective techniques, the operator’s assessment of the task difficulty can provide insight to researchers and system designers. Tsang and Wilson (1997) noted that the widespread use of subjective measures was due, in part, to their ease of use, operator acceptance and face validity. Two of the most often-used subjective ratings, the NASA Task Load Index (TLX; Hart & Staveland, 1988) and the Subjective Workload Assessment Technique (SWAT; Reid & Nygren, 1988), were used as dependent measures for the present study.

The NASA TLX consists of six subscales used to measure an operator’s mental workload. The subscales are Mental Demand, Physical Demand, Temporal Demand, Effort, Frustration, and Performance. The subscales are ordered from Low to High, except for the Performance subscale, which has a Good to Poor order. Each subscale is divided into 20 intervals with a bipolar descriptor on each end (i.e., Low/High). The rating instrument is normally given after each scenario or at targeted points in the session whereby the operator is asked to select one of the intervals for each subscale.

For the present study, a TLX-M4 overall mental workload scale was calculated using four of the six TLX dimensions. For the TLX-M4 analysis, the weights and composite scores from the mental demand, temporal demand, effort, and frustration subscales were used. The
Performance subscale was removed from the TLX-M4 analysis due to the fact that it is a personal judgment of primary task performance, which was already used as a main indicator of mental workload. The Physical Demand subscale was removed from the TLX-M4 analysis because this supervisory task does not require any joystick or throttle control, or other physically demanding actions. This was reflected in the TLX Sources of Workload scaling procedure, as 10 of the 12 participants gave the Physical Demand factor a zero overall weight.

A scaling procedure is used to combine the subscales and obtain an overall workload index. During this “sources of workload” scaling procedure, operators make a pair-wise comparison of each subscale and choose which one is more important to them for the task domain. The subscale weighting comparisons were completed once per operator, after the operator had been introduced to the task. To calculate the TLX-M4 composite scores, the relative frequency that a subscale was chosen as more important in the 6 comparisons was tallied and used as the subscale weight. These tallies ranged from zero (weight = .000, not an important factor in the task) to three (weight = .5, more important than any other factor). After each scenario, the rating for a subscale (i.e., some value 0-100) was multiplied by its particular subscale weight (i.e., 0-3), summed with the other three subscale products, and then divided by 6 (i.e., the number of pair-wise comparisons) to derive a composite score for that particular scenario. For example, in scenario # 18, one operator gave marks that valued 95, 100, 95, and 25, for the Mental Demand, Temporal Demand, Effort, and Frustration subscales, respectively. Using that operator’s given TLX-M4 subscale weights (2, 3, 1, 0), this equated to a composite score of 97.5 \( \{(95\times2 + 100\times3 + 95\times1 + 25\times0)/6\} \). A further description of each TLX subscale is provided in Appendix A.
The SWAT technique was another dependent measure used for this study. Based upon their literature review of the common factors that affect mental workload, Reid and Nygren (1988) divided SWAT into three subscales: (a) time load, (b) mental effort load, and (c) psychological stress load. Time load refers to a situation whereby an operator may be required to accomplish more tasks (either serially or simultaneously) than the given time period will allow. Mental effort load refers to the finite capacity an operator has to perform certain tasks, such as decision-making, calculating, attending to displays or text, etc. Psychological stress load refers to the general concepts that contribute to various physical or mental stressors, such as anxiety, frustration, temperature, noise, and confusion, among others. Each subscale has three descriptive levels. For example, at the end of scenario #18, the most difficult UCAV scenario, most of the operators provided an event rating of 3-3-3. This indicated a high level of all three dimensions. One operator provided an event rating of 3-3-2 for the most difficult scenario, which indicated a high level of the Time Load and Mental Effort Load dimensions and a middle level of the Psychological Stress Load dimension.

The operators also performed a scaling operation to combine the SWAT subscales into an overall workload index. Operators sorted 27 cards, consisting of all combinations of the three subscales, from lowest to highest. Conjoint measurement was used to derive the operator’s numeric scale values for each level of each subscale (Reid & Nygren, 1988). The numeric scale value (i.e., 0.0, 15.4, etc.) was assigned to each of the 27 possible levels of the three subscale dimensions (i.e., 1-1-1, 1-1-2, etc.). For example, each time an operator gave a scenario an event rating of 1-1-2, a scale value of 15.4 was assigned. Unlike the TLX procedure, the SWAT procedure does not assume that the operator-selected subscale levels are linearly related to the underlying mental workload.
The objective of the weighting procedures used by both the TLX (i.e., pair-wise comparisons) and SWAT (i.e., card-sorting procedure) methods is to get all of the subjects to share the same meaning for each subjective workload rating. A further description of each SWAT subscale is provided in Appendix B. Appendixes A and B were provided to the operators during the course of the training and data collection as reminders of the meanings of the workload dimensions for each method.

*Mental Workload Redline.* Mental workload redline refers to a point at which operator performance starts to degrade. Using the analogy of a tachometer that corresponds to an operator’s degree of mental workload, Colle and Reid (2005) stated that a redline could be thought of as the point at which the factors that are combining to overload the operator have reached a critical level. They equated this redline to the transitional area between Regions A and B as described by O’Donnell and Eggemeier (1986). These regions are graphically represented in Figure 1. This redline indicates that performance has begun to show decrements and that the mental workload levels should be addressed. A mental workload redline has been defined both in terms of primary task performance and by using the SWAT subjective measure. A performance redline was used to estimate the critical transition point between Regions A and B and it was related to the SWAT scale. A SWAT score of 40 ±10 has been estimated previously as a redline value (Colle & Reid, 2005; Reid & Colle, 1988). A redline score has not yet been estimated for the NASA TLX scale.

*Secondary Task Performance.* Often, secondary task measures are used to ascertain the point at which an operator’s primary task performance in a multi-task environment begins to degrade. O’Donnell and Eggemeier (1986) described three major categories of secondary task paradigms: (a) task loading, (b) subsidiary tasks and (c) performance operating characteristics
(POC). In the task loading paradigm, operators are instructed to maintain performance on the secondary task, even if decrements to the primary task result. In the subsidiary task paradigm, operators are instructed to maintain the primary task performance to the best of their ability, at the expense of completing the secondary task. In the subsidiary task paradigm, it is assumed that as primary task demands increase, requiring more cognitive effort, performance in the secondary task suffers. Some examples of simple laboratory secondary tasks include mental arithmetic calculations and memory recall exercises. In the POC paradigm, emphasis is not just on completing the primary task. Performance on both the primary and secondary tasks is plotted as the emphasis of each task is varied. All of these paradigms require procedural constraints and special assumptions, but they have been useful theoretically and practically. However, secondary tasks can intrude on primary task performance and may reduce the operator’s acceptance and sense of realism. Given that the specific primary task demands of the UCAV operator’s mission are not yet thoroughly understood, secondary task techniques were not used in the present study.
III. Goals of the Study

The present study had several goals. Primarily, it investigated where performance starts to degrade as a function of the number of UCAVs, vehicle airspeed, and difficulty level. The point at which these operationally relevant parameters combined to overload the operator was reflected in the task performance measures. Identifying a performance redline should provide guidelines for implementing operator requirements for unaided UCAV systems. Understanding these limitations should contribute to a more effective operational weapons system.

The subjective measures, TLX-M4 and SWAT, were used to evaluate the levels of subjective workload engendered by the parameters of the study. These measures were used as converging operations to estimate mental workload redlines for the task parameters. Another goal of the present study was to test whether the SWAT redline estimate of (40 ±10) will be confirmed in the UCAV domain using the performance redline. Previous performance-based research (Reid & Colle, 1988; Colle & Reid, 2005) indicated that SWAT values of 40 ±10 signaled a range for establishing that a degree of mental workload may be nearing a critical level. This redline estimate could be useful to system designers as remotely operated vehicles continue to be developed.

Finally, the present study sought to develop a TLX-M4 redline rating comparable to the SWAT redline estimate of 40 ±10. Because both methods are very popular in mental workload-related research, any equivalencies that could be determined would help to show agreement between the individual techniques. While several authors have discussed these two techniques in general terms (see, for example, Colle & Reid, 1998; Moroney, Biers, & Eggemeier, 1995; Tsang & Wilson, 1997; Annett, 2002; Rubio, Díaz, Martín, & Puente, 2004), none explored the connection between the ratings in the context of a performance redline. The TLX instrument
contains two subscales, performance and physical workload, which are not directly related to mental workload. Given that the performance and physical workload subscales may not effectively contribute to the redline evaluation, the data analyses were conducted with (i.e., All six TLX factors) and without them (i.e., TLX-M4). Appendix C shows the weighted TLX-M4 scale.
IV. Method

a. Participants

Twelve participants who completed the study ranged between 19-39 years of age ($M = 24.5$, $SD = 5.66$). There were 11 males and 1 female in the study. Six were USAF company grade officers and six were employees of a subject pool here of AFRL. The officers were not paid for their involvement, but the six subject pool members were paid their standard subject pool-employee wage ($M = \$17.53/\text{hour}$, $SD = \$2.56$). All 12 participants indicated that they predominantly use their right hand for manipulating the computer mouse and each signed an informed consent document. Two other participants were dropped during their training portion. One participant was dropped because she consistently failed to show up for training sessions and for poor progress in training. The second participant was dropped after I discovered that he was slightly color deficient, which negatively affected his performance. Two satisfactory replacements were found who completed the study. All 12 participants were screened for normal or corrected-to-normal visual acuity. The total time for training and data collection was 136.5 hours ($M = 11.38$ hours/participant, $SD = 2.7$). Training took approximately 7 hours for each participant. Data collection took approximately 4 hours for each participant and was divided into 2-3 sessions.

b. Facilities, Software and Apparatus

The facilities of the Air Force Research Laboratory, Flight Psychophysiology Laboratory (FPL) (http://www.hec.afrl.af.mil/Organization/HECP/fpl.asp) at Wright-Patterson Air Force Base were used. The H$_2$O UCAV simulation software, developed by the Boeing Company, was the task environment used by the operators. The Boeing Company provided this simulation software to the FPL for research use as a result of their collaboration on the Augmented
Cognition program. The H$_2$O simulator runs on a Dell Pentium 4 computer with a 3.2 GHz processor, with a keyboard and mouse, and is displayed on a Sharp LC-M3700 monitor, 95 cm wide $\times$ 57 cm high $\times$ 10 cm deep, with a screen resolution of 2560 $\times$ 768 pixels. The mouse served as the primary input device.

\textit{c. Scenario}

The H$_2$O simulator presented the operators with simulated mission scenarios in which they supervised varying numbers of UCAVs in sets of four, known as strike packages. The number of strike packages was 1, 2, and 3, totaling four, eight, or twelve UCAVs in a given scenario. The operators directed the UCAVs to engage the ground target icons by creating a UCAV-to-target pairing. The target icons represented a section of ground area. These ground areas contained the objects of interest (i.e., valid targets and distracters).

The monitor served as a tactical situation display (TSD). This TSD, partially shown in Figure 2, provided a view of the vehicles and their predetermined paths, laid over a background topographical map. The TSD displayed a red line that indicated the boundary of area of responsibility (AOR), in which the UCAVs were required to engage the target icons (target icon #100 is enclosed by a rectangle in Figure 2).
The UCAVs traveled from the bottom of the screen to the top during each scenario. While there was no dynamic route re-planning capability, each UCAV did automatically divert slightly from its original course to approach the target icon at the proper angle during an attack sequence. After delivering its munitions, the UCAV then rejoined its original flight path at the most convenient point.

The operators were responsible for pairing the appropriate vehicle to the appropriate target icon, based on the munitions required for that target (i.e., small-, medium-, and large-diameter). The pairings were made by left-clicking on the appropriate munition “token” adjacent to the vehicle of interest, and then left-clicking on the selected target icon. As illustrated in Figure 3, a dashed line was then displayed between the two objects, with a distinctive color.
scheme representing the exact function that was selected (i.e., Capture Synthetic Aperture Radar (SAR) Image, Direct Attack, or Capture SAR/Direct Attack).

![Figure 3. Each UCAV is engaging a target icon.](image)

The operator was not allowed to create a UCAV-to-target pairing that required the vehicle to go backwards. This situation was explicitly described as violating one of the rules of engagement (ROE). A straight lateral (horizontal) attack path was acceptable, given that the operator adhered to the other constraints and ROE, which are described below.

Each SAR image, regardless of the level of attack difficulty, contained six valid target vehicles. The low attack difficulty SARs contained only valid targets, while the high attack difficulty SARs contained valid targets along with a number of distracters and trees. These valid targets and distracters are shown in Figures 5 and 6, respectively. Different combinations of the six valid targets, with at least one of each type, were placed in the high-difficulty SAR images (i.e., two C-targets, three B-targets, and one A-target), such that the operator did not know beforehand exactly how many of each type of valid SAR targets were present. This necessitated a thorough visual search of each SAR image. The target icons were placed in a corridor for each
strike package. These corridors had a consistent threat density (e.g., number of target icons) for each strike package throughout the scenario. In other words, the target icons were evenly spaced along the vertical track of the corridor so that the workload was constant throughout the mission.

Figure 4. Examples of valid targets that were used in the simulated SAR images. “A” targets required a large-diameter munition, “B” targets required a medium-diameter munition, and “C” targets required a small-diameter munition.
Synthetic aperture radar (SAR) technology was developed in the late 1970s by NASA. A radar antenna transmits and receives radar pulses to obtain an image. The reflected energy arrives at the receiver at different times, calculated to be representative of the distance traveled. These calculations are used to create a topographic map of the target area, to include objects in that area. A shadow is created by the reflected energy, representative of the objects’ physical characteristics, such as height, width, length, and orientation. Figures 6 and 7 show simulated examples of these SAR images.

The valid SAR targets varied in type, such that an appropriately-sized munition was required to achieve a satisfactory hit. Additionally, the DMPI must have been placed within a defined distance of the SAR target, known as the “blast radius”, to achieve a satisfactory hit. For
example, all six SAR targets shown in Figure 6 required a small-diameter munition. The superimposed circle located below each SAR target represents the size of the blast radius. This was the effective radial miss distance for all three munition types (small-, medium-, and large-diameter). The blast radius equated to 25 pixels. The operators were trained how to recognize the point to which the blast radius extended for the munitions.

All three types of valid SAR targets are shown in Figure 7, along with various distracter objects. The SAR target surrounded by the circle was categorized as a “C” target and required a small-diameter munition. The SAR targets surrounded by a rectangle were categorized as “B” targets and required a medium-diameter munition. The SAR targets surrounded by a diamond were categorized as “A” targets and required a large-diameter munition. The remaining vehicles were categorized as distracters. The simulated SAR image shown in Figure 7 required multiple UCAVs (usually two) to engage it to account for all three types of target vehicles. Each UCAV was capable of carrying only two types of munitions at once (i.e., small/medium, small/large, medium/large).
Figure 6. A typical easy SAR image. The six circles positioned just below each SAR target vehicle were added to identify the vehicles and to illustrate the effective blast radius for all three munition types. The circles were not part of the actual display.
In the low attack difficulty condition, the SAR images contained only C-targets. Only one UCAV was required to engage these target icons. During the high attack difficulty condition, the SAR images contained A-, B-, and C-targets, and required coordinated attacks. That is, if the UCAV was carrying only small- and large-diameter munitions, a second UCAV carrying medium-diameter munitions was needed to engage the target icon to completely process the SAR image (i.e., assign proper munitions to all three valid target types). In Figure 2, the rectangle surrounds a target icon. The text adjacent to the target icon depicted the type of SAR target(s) contained in the SAR image (i.e., SA-6 = All Cs, low attack difficulty; SA-8 = As, Bs,
and Cs, high attack difficulty), and the target number, which was used to assign a “corridor” of targets to a particular strike package. As can be seen in Figure 2, the UCAVs from one corridor (i.e., UCAVs #11-14) were not allowed to prosecute targets icons located in a different corridor (i.e., target icons #130, 134, etc.).

The short, rectangular area along the track path adjacent to the nose of UCAV #14 on the left side of Figure 3 provided the operators with a visual indication of when the vehicle completed the SAR download process. On the right side of Figure 3, UCAV #13 has downloaded the SAR image and is ready to view the target area. Note the absence of the small, rectangular area previously adjacent to the nose of the UCAV, as well as the change in the symbology of the target icon. The small square roughly halfway between the UCAV and the target icon represented the weapons release point. This square indicated to the operator the point by which the munitions must be cleared for release.

Clearing the weapons for release required the operator to initiate the SAR image download function, open the SAR target window, correctly place the six designated mean points of impact (DMPIs) using the appropriate munitions, and then click the “OK”, “Finish”, or “Apply” button. If the SAR was not processed by the time the vehicle reached this critical point, the UCAV had overflown the target and was not allowed to prosecute it. It could be prosecuted by a trailing UCAV, if that UCAV was carrying the appropriately-sized munitions. Operators were instructed to use this course of action to correctly process as many SARs as possible during each scenario.

After using the tokens to complete either a Capture SAR function or a SAR/Direct Attack function, the target icon symbol changed in appearance to indicate that the SAR image might then be viewed. In the left side of Figure 3, the target icon has been paired to a vehicle, but has
not yet completed its download. The SAR for the target icon on the right side of Figure 3 has been downloaded and is ready to be viewed. A single left-click of the target icon enabled a pop-up targeting window showing the simulated SAR image. These simulated SAR images showed hostile and friendly forces, which were classified as valid SAR target vehicles, and invalid distracter objects, respectively. Valid targets were to be destroyed, while distracters were to be avoided. The operators were trained to discriminate between these two vehicle groups, and then assign a DMPI to each valid SAR target.

\textit{d. Experimental Design}

The present study used a $3 \times 3 \times 2$ repeated-measures factorial design. As shown in Table 1, the first column lists the arbitrary scenario numbers, which identified the scenario conditions. The second column shows one of the possible combinations of variables from the factorial manipulation. Listed first is the number of UCAVs that were supervised, next is the UCAV airspeed, followed by the level of attack difficulty. The third column lists the number of target icons that were present in the scenario, representative of the number of strike packages. The fourth column lists the number of valid SAR targets that were presented in the SAR image.
Table 1. Cells of the $3 \times 3 \times 2$ repeated-measures factorial design.

<table>
<thead>
<tr>
<th>Scenario #</th>
<th># of UCAVs/UCAV Airspeed/Difficulty</th>
<th>Total # of Target Icons</th>
<th>Total Valid SAR Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4/900/Low</td>
<td>20</td>
<td>120</td>
</tr>
<tr>
<td>2</td>
<td>4/900/High</td>
<td>20</td>
<td>120</td>
</tr>
<tr>
<td>3</td>
<td>4/1200/Low</td>
<td>20</td>
<td>120</td>
</tr>
<tr>
<td>4</td>
<td>4/1200/High</td>
<td>20</td>
<td>120</td>
</tr>
<tr>
<td>5</td>
<td>4/1500/Low</td>
<td>20</td>
<td>120</td>
</tr>
<tr>
<td>6</td>
<td>4/1500/High</td>
<td>20</td>
<td>120</td>
</tr>
<tr>
<td>7</td>
<td>8/900/Low</td>
<td>40</td>
<td>240</td>
</tr>
<tr>
<td>8</td>
<td>8/900/High</td>
<td>40</td>
<td>240</td>
</tr>
<tr>
<td>9</td>
<td>8/1200/Low</td>
<td>40</td>
<td>240</td>
</tr>
<tr>
<td>10</td>
<td>8/1200/High</td>
<td>40</td>
<td>240</td>
</tr>
<tr>
<td>11</td>
<td>8/1500/Low</td>
<td>40</td>
<td>240</td>
</tr>
<tr>
<td>12</td>
<td>8/1500/High</td>
<td>40</td>
<td>240</td>
</tr>
<tr>
<td>13</td>
<td>12/900/Low</td>
<td>60</td>
<td>360</td>
</tr>
<tr>
<td>14</td>
<td>12/900/High</td>
<td>60</td>
<td>360</td>
</tr>
<tr>
<td>15</td>
<td>12/1200/Low</td>
<td>60</td>
<td>360</td>
</tr>
<tr>
<td>16</td>
<td>12/1200/High</td>
<td>60</td>
<td>360</td>
</tr>
<tr>
<td>17</td>
<td>12/1500/Low</td>
<td>60</td>
<td>360</td>
</tr>
<tr>
<td>18</td>
<td>12/1500/High</td>
<td>60</td>
<td>360</td>
</tr>
</tbody>
</table>
The number of UCAVs supervised was 4, 8, or 12. The UCAV vehicle airspeeds were preset to 900 knots, 1200 knots and 1500 knots, and did not vary during a scenario. These parameters were selected based on previous research. As depicted in Figures 6 and 7, there were two levels of attack difficulty, low and high, respectively. Several factors determined the level of attack difficulty. Those factors included the number of vehicles required to successfully prosecute a SAR image (usually two), the use of a single valid target type (low difficulty condition) or several valid target types (high difficulty condition) within the SAR image, and the absence (low difficulty) or presence (high difficulty) of distracter objects within the SAR image.

In the low-difficulty condition, one UCAV engaged a target icon. There were six “C” targets in each SAR image, as shown in Figure 6 (positioned just above the circles). In the high-difficulty condition, two, or possibly three, UCAVs were required to engage the target icon. Seen at the top left of the SAR window in Figure 7, two UCAVs (#11 and #14) are available to assign munitions to the various SAR targets. The “C” target vehicle that is surrounded by the oval was to be prosecuted using the small-diameter munition, carried by UCAV #11. The three “B” vehicles that are surrounded by the rectangle were to be prosecuted using medium-diameter munitions, carried by UCAV #14. The two “A” vehicles surrounded by the diamonds were to be prosecuted using large-diameter munitions, carried by both UCAV #11 and #14. By varying the combinations of SAR target vehicles, the operator was required to visually search each SAR image. Each operator was trained to use a systematic (i.e., counter-clockwise) scan pattern, to reduce the time wasted by randomly searching, and to maintain consistency across operators.

The 3 × 3 × 2 design resulted in 18 test scenarios. Table 2 shows the randomized order in which the operators received the 18 scenarios. The cells of the table show the scenario number. The operators were tested in three blocks of six trials. The six possible block orders were used
equally often, with two operators per order. Operators were randomly assigned to a block order. The number of UCAVs being supervised was constant for each block. Within a block, the operators were tested on all six possible combinations of airspeed and difficulty. The order of testing for each of these six combinations was determined by random sampling without replacement.

Table 2. Randomized order of operator trials. Scenario numbers are listed in the cells of the table.

<table>
<thead>
<tr>
<th>Operator Number</th>
<th>Trial #</th>
<th>Block</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>12</td>
<td>1</td>
<td>15</td>
<td>12</td>
<td>18</td>
<td>4</td>
<td>12</td>
<td>15</td>
<td>16</td>
<td>9</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>10</td>
<td>3</td>
<td>13</td>
<td>11</td>
<td>15</td>
<td>1</td>
<td>10</td>
<td>14</td>
<td>13</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>9</td>
<td>5</td>
<td>14</td>
<td>8</td>
<td>14</td>
<td>5</td>
<td>7</td>
<td>13</td>
<td>18</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>7</td>
<td>2</td>
<td>17</td>
<td>9</td>
<td>16</td>
<td>2</td>
<td>11</td>
<td>17</td>
<td>17</td>
<td>12</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>8</td>
<td>4</td>
<td>18</td>
<td>10</td>
<td>17</td>
<td>6</td>
<td>9</td>
<td>16</td>
<td>14</td>
<td>11</td>
<td>3</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td>11</td>
<td>6</td>
<td>16</td>
<td>7</td>
<td>13</td>
<td>3</td>
<td>8</td>
<td>18</td>
<td>15</td>
<td>10</td>
<td>5</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td>13</td>
<td>10</td>
<td>12</td>
<td>5</td>
<td>13</td>
<td>17</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>18</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>B</td>
<td>17</td>
<td>9</td>
<td>7</td>
<td>1</td>
<td>6</td>
<td>17</td>
<td>13</td>
<td>7</td>
<td>6</td>
<td>1</td>
<td>16</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>B</td>
<td>18</td>
<td>11</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>15</td>
<td>16</td>
<td>12</td>
<td>2</td>
<td>3</td>
<td>15</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>B</td>
<td>16</td>
<td>7</td>
<td>11</td>
<td>4</td>
<td>5</td>
<td>18</td>
<td>14</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>13</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>B</td>
<td>14</td>
<td>12</td>
<td>9</td>
<td>6</td>
<td>3</td>
<td>16</td>
<td>18</td>
<td>8</td>
<td>3</td>
<td>4</td>
<td>14</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>B</td>
<td>15</td>
<td>8</td>
<td>10</td>
<td>2</td>
<td>4</td>
<td>14</td>
<td>15</td>
<td>11</td>
<td>4</td>
<td>5</td>
<td>17</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>C</td>
<td>5</td>
<td>13</td>
<td>2</td>
<td>14</td>
<td>8</td>
<td>9</td>
<td>1</td>
<td>4</td>
<td>11</td>
<td>16</td>
<td>9</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>C</td>
<td>1</td>
<td>7</td>
<td>6</td>
<td>18</td>
<td>9</td>
<td>11</td>
<td>3</td>
<td>2</td>
<td>12</td>
<td>17</td>
<td>12</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>C</td>
<td>6</td>
<td>16</td>
<td>4</td>
<td>17</td>
<td>7</td>
<td>12</td>
<td>4</td>
<td>3</td>
<td>9</td>
<td>13</td>
<td>10</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>C</td>
<td>2</td>
<td>18</td>
<td>5</td>
<td>16</td>
<td>12</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>10</td>
<td>18</td>
<td>11</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>C</td>
<td>3</td>
<td>14</td>
<td>1</td>
<td>13</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>15</td>
<td>7</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>C</td>
<td>4</td>
<td>15</td>
<td>3</td>
<td>15</td>
<td>11</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>14</td>
<td>8</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

Block order: BCA ABC CBA BAC CAB ACB BCA CBA CAB BAC ACB ABC
Scale order: ST TS ST TS TS ST TS ST TS ST ST TS

Note. For the block order, A = 4 UCAVs, B = 8 UCAVs, and C = 12 UCAVs. For the scale order, T = TLX and S = SWAT.

e. Procedure

Training. One-on-one training sessions were conducted to ensure consistent and stable performance. An initial session was used to introduce the subjects to the task and familiarize
them with the subjective measures. Each operator was introduced to the H$_2$O simulator using a single strike package in a low-difficulty scenario while flying at a slower airspeed (i.e., 800 knots) than the slowest one used in the actual experiment. The training scenarios progressed in difficulty as the operators’ proficiency warranted. The operators were then trained to stable performance. This took between 3-5 sessions, with each training session lasting between 60-90 minutes. Additionally, the operators completed the TLX sources of workload scaling procedure, as well as the SWAT card sorting procedure as described by Reid, Potter and Bressler (1989). Ten of the twelve participants had previous experience with UCAV simulators. Seven of those ten had previously served as participants in UCAV studies in our laboratory, with an earlier simulation environment. Three of those ten had served in a separate laboratory, using a slightly different simulation environment with a similar SEAD mission task.

For the final three training scenarios, and for every task scenario, each operator completed the TLX rating and the SWAT event scoring for that individual scenario, in a counterbalanced order. These final three training scenarios allowed the operators to become familiar with the rating instruments and have an opportunity to ask questions about the relationship of the subscales to the task. Furthermore, the final three training scenarios were used to introduce the operators to scenarios that reflected the highest and lowest degrees of difficulty, as well as a middle degree of difficulty. The operators were then instructed to use the full range of subjective ratings, by considering the scenarios with the lowest and highest difficulty levels as points of reference. Introducing the operators to the spectrum of task difficulty minimized the contextual effects of the operators using only limited ranges of the rating scales, as described by Colle and Reid (1998). This effect might have occurred if the operators were unaware of the low and high endpoints of the task difficulty. The training
process ended when the operators were able to successfully complete two scenarios of low difficulty without error (i.e., one strike package flying at the lowest airspeed, using the single UCAV attack procedure). A step-by-step description of the training procedure is provided in Appendix D.

Task. The H2O UCAV simulator let operators perform simulated SEAD missions while controlling multiple UCAVs. The operator created as many UCAV-to-target pairings as possible during each scenario and processed the subsequent SAR images. Each scenario started out with the vehicles pre-positioned at the bottom of the TSD and pointing upward. There was an interactive zoom function that allowed the operator to change the amount of the TSD that is viewable. At the standard zoom setting, the first few target icons were initially in view. The operators were trained to keep the use of this zoom function to a minimum. Because the training sessions occurred on different dates than the data collection sessions, the operators were each given one practice trial using the easiest scenario parameters (4/900/low) before the test scenarios began.

To initiate a scenario, the operator clicked the “Start” button at the top of the TSD, which began the UCAVs’ travel on their predetermined routes and the system clock timer. The operator then began to make the appropriate UCAV-target pairings. As shown in Figure 3, the symbology of the target icons updated according to the function that was initiated. As the UCAVs traveled upward, the target icons indicated to the operator that the SAR download process was complete and available for viewing. The operators used the mouse to left-click on the target icon and bring up the SAR targeting window. The SAR targeting window, similar to Figures 6 and 7, then appeared on the right side of the TSD.
The operators were required to correctly place six DMPIs per SAR, on the appropriate vehicles, inside of the defined blast radius to receive proper credit. Operator training emphasized that all six DMPIs be placed before closing the SAR targeting window and moving on to the next sub-task. However, operator training also emphasized, that if the vehicle did happen to pass over the weapons release point while the DMPIs were being placed, the operator should close the SAR window and move on to the next sub-task.

After placing the six DMPIs, the operator clicked one of several rectangular buttons, as pictured in the top right of the SAR targeting window shown in Figure 6. By clicking “Apply”, the operator assigned the DMPIs to the target vehicles and, in the difficult condition, was then ready to select the other UCAV (i.e., UCAV #14) to allocate munitions being carried by the second UCAV. By clicking the square “Undo” button as shown in Figures 6 and 7, located just below the “Apply” button, the operator removed the last DMPI that was placed in the SAR image window. By clicking “Cancel”, the operator could close the SAR targeting window prior to completing the DMPI-placement process. This would disengage the UCAV from the target icon and remove any DMPIs that had been placed.

This event might occur if an operator inadvertently paired a UCAV to a target icon and then realized that the UCAV was not carrying the appropriate quantity or type of munitions required to properly engage the existing target vehicles. The SAR targeting window remained open until the operator selected “OK” or “Finish”. By clicking “OK” or “Finish”, the operator acknowledged that the SAR image had been completely processed. This act of acknowledgment cleared the weapons for release and allowed for the munitions to fly when the UCAV reached the weapons release point. The SAR window remained open until it was closed by the operator. If a UCAV passed its weapons release point before the operator cleared the weapons for release, a
brief tone signaled the operator and those munitions did not fly. The scenario ended after all authorized munitions had landed. There was a brief inter-trial interval as the two subjective measures were collected, and as the simulation software was loaded with the next scenario. During the data collection trials, the presentation order of the subjective measures was randomized so that half of the participants received the TLX first and the other half received the SWAT first. Each participant’s data collection ended after all 18 scenarios were completed, and the associated subjective mental workload forms were completed.

\textit{f. Statistical Analyses}

Two different types of statistical analyses were conducted. The first type consisted of four $3 \times 3 \times 2$ factorial ANOVAs on the two performance dependent variables (percent kills, percent DMPIs) and the two subjective workload dependent variables (SWAT, TLX-M4). The second type consisted of accomplishment score analyses which estimated performance-based mental workload redlines. A $p$-value of 0.05 was used as a criterion for statistical significance. All $p$-values were adjusted using the Greenhouse-Geisser correction ($p_{g-g}$). The F-ratios are reported with the original degrees of freedom, but the Greenhouse-Geisser Epsilon ($\varepsilon_{g-g}$) that was used to correct the degrees of freedom is reported.

\textit{Factorial ANOVAs.} The two performance dependent variables that were analyzed were the percentage of kills and the percentage of correctly-placed DMPIs. The two subjective workload dependent variables that were analyzed were the SWAT and the TLX-M4. For performance analyses, percentage of kills was the primary criterion. The percentage of correctly-placed DMPIs was a secondary criterion. A DMPI was classified as correctly placed if the operator assigned the correct munition type to the appropriate target (i.e., medium munition on a “B” target) within the blast radius, and then authorized the munitions to fly before the
UCAV reached the weapons release point. A kill was defined as a processed SAR image that had all six DMPIs correctly placed before the UCAV reached the weapons release point for that particular target icon. In other words, if the operators correctly placed six DMPIs on the proper target vehicles and clicked the “OK” button before the UCAV reached the weapons release point, one kill was credited. Table 3 shows how the DMPIs and kills were defined and scored.

<table>
<thead>
<tr>
<th># of Correctly-Placed DMPIs/SAR</th>
<th>Operator Cleared for Release (Click “OK”)</th>
<th># Credited DMPIs</th>
<th>Kills</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Yes</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>N &lt; 6</td>
<td>Yes</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>No</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>N &lt; 6</td>
<td>No</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The second type of dependent variables that were analyzed was subjective mental workload ratings. Both subjective mental workload methods use scales that range from 0-100, with 100 representing very high mental workload. The TLX-M4 scale was used the sources of workload scaling procedure, which included only the mental demand, temporal demand, effort, and frustration subscales pair-wise comparisons. As shown in Appendix E, a single SWAT scale for the group of participants was developed using the standard card-sorting procedure as described by Reid, Potter and Bressler (1989).

**Accomplishment score.** An accomplishment score method, as described by Colle and Reid (1997, 1999, 2005) was used to compare the number of tasks (i.e., number of kills and number of correctly-placed DMPIs) successfully completed in each scenario, using a fixed time interval, nine minutes. The number of tasks completed in the nine-minute interval was used to create accomplishment scores for each trial. This interval represented the time it took an aircraft
flying at 1500 knots, the fastest airspeed, to reach the end of the scenario, to equate intervals for the other two airspeeds (i.e., 900 and 1200 knots). This established an equivalent fixed time interval, so that the number of successfully completed tasks (i.e., SARs killed, DMPIs placed) from one experimental condition could be directly compared to other experimental conditions. The analysis unit is the number of targets killed or correctly-placed DMPIs in nine minutes.

The focus of the analysis of accomplishment scores was: (a) to determine performance redlines for each difficulty level, and (b) to relate those redlines to corresponding SWAT and TLX-M4 rating levels. Accomplishment scores were analyzed by conducting separate ANOVAs on low difficulty and high difficulty. Both ANOVAs had one factor with 9 levels, created from the $3 \times 3$ factorial combinations of the airspeed and the number of vehicles. This factor was the maximum number of kills (and DMPIs) that could be reached in nine minutes. The single factor was analyzed using orthogonal polynomial contrasts with uneven spacing. The maximum possible numbers of kills and DMPIs were established by determining which SAR target icons could be reached by the 9-minute point by flying and performing optimally. A scoring marker was placed at the 9-minute point to calculate the correct numbers of kills and DMPIs, versus the maximum numbers possible. This marker was used for data analysis only and was transparent to the operators.
V. Results

Because very few distracter objects were destroyed (32 in all), the analyses of the performance data focused on the percentage of kills and correctly-placed DMPIs. Nine of the twelve participants did not hit any distracters. For each subject, there was a total of 6,974 distracters divided among the 9 high-difficulty scenarios ($M = 387$/strike package). The total number of distracters presented to all 12 subjects was 83,688. Of the three participants who did hit distracters, one hit 18, one hit 2 and one hit 12. The percentage of distracters destroyed was .04% for all participants across all trials. See Appendix F for a breakout table of the distracters that were destroyed, by subject and by scenario.

a. Performance

Two separate $3 \times 3 \times 2$ repeated-measures ANOVAs were used to analyze the performance variables, the percentage of kills and the percentage of correctly-placed DMPIs. The factors were the number of UCAVs supervised (4, 8, or 12), the UCAV airspeeds (900, 1200, or 1500 knots), and the levels of attack difficulty (low and high). Note that these analyses use the data for the entire scenario. Results were similar for data based on only the 9-minute fixed time period. See Appendix AD.

Percent Kills. As Figures 8 and 9 show, fewer kills were made for the high attack difficulty than for the low attack difficulty, $F(1, 11) = 404.6$, $MSE = 165.52$, $p < .0001$. The mean percentage of kills for the high attack difficulty was only 51.7%, compared to 86.9% for the low difficulty scenarios. There was a statistically significant main effect of the Number of Vehicles, $F(2, 22) = 94.78$, $MSE = 203.03$, $\epsilon_{g-g} = .74$, $p_{g-g} < .0001$. As Figures 8 and 9 show, the mean percentage of kills was 84.9% for 4 vehicles, 70.6% for 8 vehicles, and 52.3% for 12 vehicles. There was also a statistically significant main effect of the levels of Airspeed, $F(2, 22)$
\(= 407.94, \text{MSE} = 24.89, e_{g-g} = .78, p_{g-g} < .0001.\) The mean percentage of kills was 81.0% at 900 knots, 69.7% for the 1200 knots, and 57.2% for 1500 knots.

**Figure 8.** Percent Kills for High Attack Difficulty

**Figure 9.** Percent Kills for Low Attack Difficulty Scenarios.
As can be seen in Figures 8 and 9, the pattern of results was different for high and low
difficulty scenarios. The three-way interaction of the levels of attack difficulty, number of
vehicles, and airspeed, was statistically significant, \( F(4, 44) = 8.04, MSE = 50.69, e_{g-g} = .60, p_{g-g} = .001 \). As shown in Figure 9, the percentage of kills is at ceiling for four vehicles with low
attack difficulty for all airspeeds. The percentage of kills is also at or near ceiling for eight
vehicles for the two lower airspeeds. Consistent performance degradations with increasing
airspeeds are seen only when twelve vehicles were being supervised. In contrast, as can be seen
in Figure 8, when targets were difficult to detect and attack, both airspeed and number of
vehicles had major effects at all levels. The three two-way interactions also were statistically
significant. The complete ANOVA table is listed in Appendix G.

Follow-up ANOVAs were consistent with the above description. A 3 × 3 follow-up,
repeated-measures ANOVA on the high attack difficulty conditions only was also conducted.
Consistent with the near-parallel lines shown in Figure 8, the two-way interaction of airspeed
and number of vehicles supervised was not statistically significant, \( F(4, 44) = 2.71, MSE = 57.71, e_{g-g} = .65, p_{g-g} = .07 \). The main effects of both airspeed and number of vehicles supervised
were statistically significant, \( F(2, 22) = 106.55, MSE = 79.78, e_{g-g} = .91, p_{g-g} < .0001 \), and \( F(2, 22) = 115.45, MSE = 110.01, e_{g-g} = .92, p_{g-g} < .0001 \), respectively. The complete ANOVA table
is listed in Appendix H. A 3 × 3 follow-up, repeated-measures ANOVA on the low attack
difficulty conditions only was also conducted. In contrast to the high attack difficulty ANOVA,
this one produced a different pattern. Is this case, the two-way interaction of airspeed and
number of vehicles supervised was statistically significant, \( F(4, 44) = 16.51, MSE = 40.48, e_{g-g} = .67, p_{g-g} < .0001 \). The two main effects were also statistically significant. The complete
ANOVA table is listed in Appendix I.
Percent DMPIs Correctly Placed. As expected, the percent of correctly-placed DMPIs showed similar patterns to the percent of kills. A correctly-placed DMPI was credited when the operator placed the correct munition type on the appropriate target (i.e., medium munition on a “B” target) within the blast radius, with the operator authorizing the munitions to fly before the UCAV reached the weapons release point.

As Figures 10 and 11 show, fewer DMPIs were placed for the high level of attack difficulty than for the low level of attack difficulty, \( F(1, 11) = 659.97, \text{MSE} = 71.07, p < .0001. \) The mean percentage of DMPIs for the low level of attack difficulty was 91.0%, as compared to 61.5% for the high level of attack difficulty. As Figures 10 and 11 also show, performance decreased when more vehicles were supervised. There was a statistically significant main effect of the Number of Vehicles, \( F(2, 22) = 164.61, \text{MSE} = 98.91, \epsilon_{g-g} = .83, p_{g-g} < .0001. \) The mean percentage of DMPIs was 90.5% for 4 vehicles, 77.7% for 8 vehicles, and 60.6% for 12 vehicles. There was a statistically significant main effect of levels of Airspeed, \( F(2, 22) = 637.16, \text{MSE} = 15.54, \epsilon_{g-g} = .98, p_{g-g} < .0001. \) As Figures 10 and 11 show, the mean percentage of correctly-placed DMPIs was 87.6% at 900 knots, 77.0% at 1200 knots, and 64.2% at 1500 knots.

There was a statistically significant three-way interaction of the levels of Difficulty by Number of Vehicles by Airspeed, \( F(4, 44) = 14.89, \text{MSE} = 29.67, \epsilon_{g-g} = .68, p_{g-g} < .0001. \) All other two-way interactions were also statistically significant. The pattern of results, as shown in Figures 10 and 11, was similar to the pattern shown in Figures 8 and 9 for percent kills. In the low level of attack difficulty, shown in Figure 11, the percent of correctly-placed DMPIs was at or near ceiling for four vehicles across all airspeeds, as well as the two lower airspeeds for eight vehicles. Performance dropped off consistently only when twelve vehicles were flown at the highest airspeed. Conversely, Figure 10 shows that when the operators had trouble
discriminating between the valid targets and the distracter objects, the percentage of correctly-placed DMPIs decreased for all conditions. See Appendix J for the complete ANOVA table.

A follow-up ANOVA of the high level of attack difficulty showed major effects at all levels of both airspeed and number of vehicles supervised. However, unlike the two-way interaction for the percentage of kills for the high level of attack difficulty, the two-way interaction for the percentage of correctly-placed DMPIs for the high level of attack difficulty was statistically significant, $F(4, 44) = 6.59$, $MSE = 34.85$, $\varepsilon_{g-g} = .69$, $p_{g-g} = .001$. As can be seen in Figure 10, performance appears to drop off more rapidly with increasing airspeed for eight vehicles, than for four or twelve. The main effects of both Airspeed and Number of Vehicles supervised were statistically significant, $F(2, 22) = 220.38$, $MSE = 65.14$, $\varepsilon_{g-g} = .88$, $p_{g-g} < .0001$, and $F(2, 22) = 234.35$, $MSE = 37.7$, $\varepsilon_{g-g} = .71$, $p_{g-g} < .0001$, respectively. The complete ANOVA table is in Appendix K.

A follow-up factorial ANOVA of only the low level of attack difficulty, as shown in Figure 11, produced results similar to those found for percent kills. The two-way interaction of airspeed and number of vehicles supervised was statistically significant, $F(4, 44) = 28.5$, $MSE = 26.76$, $\varepsilon_{g-g} = .57$, $p_{g-g} < .0001$. Both main effects also were statistically significant. The complete ANOVA table is listed in Appendix L.
Figure 10. The percentage of correctly-placed DMPIs for High Attack Difficulty Scenarios.

Figure 11. The percentage of correctly-placed DMPIs for Low Attack Difficulty Scenarios.
b. Subjective Ratings

Two separate $3 \times 3 \times 2$ repeated-measures ANOVAs were used to analyze the subjective ratings, the SWAT and the TLX-M4. The factors were the number of UCAVs supervised (4, 8, or 12), the UCAV airspeeds (900, 1200, or 1500 knots), and the levels of attack difficulty (low and high).

**SWAT ratings.** SWAT ratings can range from 0 for the lowest mental workload to 100 for the highest mental workload for rescaled values, based on the card sorts and the scaling process. Results of axiom tests are summarized in Appendix AB. As Figures 12 and 13 show, the mean SWAT ratings were higher for the high level of attack difficulty than for the low level of attack difficulty, $F(1, 11) = 82.86, MSE = 499.30, p < .0001$. The mean SWAT rating for the low level of attack difficulty was 26.9, as compared to 54.6 for the high level of attack difficulty. As Figures 12 and 13 also show, SWAT ratings increased as the number of vehicles supervised increased. There was a statistically significant main effect of the Number of Vehicles, $F(2, 22) = 137.37, MSE = 273.99, e_{g.g} = .91, p_{g.g} < .0001$. The mean SWAT rating was 17.5 for 4 vehicles, 41.4 for 8 vehicles, and 63.2 for 12 vehicles. Figures 12 and 13 also show the effect of Airspeed, $F(2, 22) = 78.63, MSE = 170.31, e_{g.g} = .78, p_{g.g} < .0001$. The mean SWAT rating was 27.5 at 900 knots, 39.9 at 1200 knots, and 54.86 at 1500 knots. However, unlike the performance data, the three-way interaction of Difficulty by Number of Vehicles by Airspeed, was not statistically significant, $F(4, 44) = 1.82, MSE = 211.17, e_{g.g} = .68, p_{g.g} = .17$. The two-way interactions also were not statistically significant. See Appendix M for the complete ANOVA table.
Figure 12. SWAT Ratings for High Attack Difficulty Scenarios.

Figure 13. SWAT Ratings for Low Attack Difficulty Scenarios.
**TLX-M4 Ratings.** Two analyses were performed on the TLX ratings. The primary analysis used only the four subscales most related to mental workload: Mental Demand, Temporal Demand, Effort, and Frustration. The secondary analysis used all six TLX subscales. The primary analysis was the TLX-M4 and those results are described below. The results for the analysis done on all six TLX subscales are listed in Appendix N.

Similar to the mean SWAT ratings, Figures 14 and 15 show that the mean TLX-M4 ratings were higher for the high level of attack difficulty than for the low level of attack difficulty, $F(1, 11) = 60.93, MSE = 419.95, p < .0001$. The mean TLX-M4 rating for the low level of attack difficulty was 42.3, as compared to 64.1 for the high level of attack difficulty. Figures 14 and 15 also show that there was a statistically significant main effect of the Number of Vehicles, $F(2, 22) = 163.27, MSE = 178.54, \epsilon_{gg} = .84, p_{gg} < .0001$. The mean TLX-M4 rating was 33.5 for 4 vehicles, 52.5 for 8 vehicles, and 73.7 for 12 vehicles. As Figures 14 and 15 also show, there was a statistically significant main effect for the levels of Airspeed, $F(2, 22) = 84.43, MSE = 112.06, \epsilon_{gg} = .78, p_{gg} < .0001$. The mean TLX-M4 rating was 42.2 at 900 knots, 52.4 at 1200 knots, and 65.1 at 1500 knots. Just as with the SWAT analysis, the three-way interaction of Difficulty by Number of Vehicles by Airspeed did not produce a statistically significant result, $F(4, 44) = 1.98, MSE = 119.87, \epsilon_{gg} = .53, p_{gg} = .16$. See Appendix O for the complete ANOVA table.
Figure 14. TLX-M4 Ratings for High Attack Difficulty Scenarios.

Figure 15. TLX-M4 Ratings for Low Attack Difficulty Scenarios.
c. Estimation of Maximum Performance: Performance Redlines

An accomplishment score analysis was performed to estimate the maximum performance, the performance redline. The performance redline estimate identifies the region above which performance does not improve. The estimate is obtained by plotting the number of accomplished acts (kills, DMPIs) against the maximum possible number in a fixed interval, in this case nine minutes. The redline is an estimate of the maximum of this curve. The accomplishment theory of average mental workload does not specify a mathematic function for this curve. Colle and Reid (2005) found that a quadratic function fit their data reasonably well and they algebraically estimated its maximum as the redline point. However, in other data sets (e.g., Colle & Reid, 1998, 2006) piecewise linear functions appeared to fit the data well. Typically, piecewise linear functions would imply that terms above the quadratic would be statistically significant in a polynomial ANOVA.

An accomplishment score analysis requires that a common fixed interval be specified, and that the mean number of accomplishment acts (kills, DMPIs) completed in that interval be obtained along with the maximum possible number of accomplishment acts. Conditions are ordered so that the maximum possible number of accomplishment acts increases. High and low attack difficulty scenarios entail qualitatively different judgments, so data from both high and low attack difficulty scenarios can not be ordered together a single axis. Therefore, high and low attack difficulty data were analyzed separately.

The conditions in which the UCAVs were flying at the fastest experimental airspeed (i.e., 1500 knots) resulted in scenario durations of nine minutes. This 9-minute period was then applied to the scenarios in which the UCAVs flew at the two remaining airspeeds (i.e., 900 and 1200 knots) so that an equivalent time interval could be established across all three airspeed
levels. The maximum possible numbers of kills and DMPIs were established by determining the number of SAR target icons could be reached by the 9-minute point while flying and performing optimally.

Table 4, modified from Table 1, lists the scenarios that were used in the present study, but in a re-ordered sequence to illustrate the order used for the accomplishment score analyses. The first column in Table 4 lists the sequence of scenarios that were used for the accomplishment score analyses, with the high attack difficulty in the top half. The second column lists the parameters of the conditions used during the scenarios. The third column lists the maximum possible number of kills that were available to the operators during each particular scenario, sequenced in a lowest-to-highest order. The fourth column lists the maximum possible number of DMPIs that were available. The maximum possible number of DMPIs are just the number of kills times six.
Table 4. Sequence of Scenarios and Maximum Possible Numbers of Kills and DMPIs used for the Accomplishment Score Analyses.

<table>
<thead>
<tr>
<th>Scenario #</th>
<th># of UCAVs / UCAV Airspeed/Difficulty</th>
<th>Maximum Possible Kills</th>
<th>DMPIs</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4/900/High</td>
<td>8</td>
<td>48</td>
</tr>
<tr>
<td>4</td>
<td>4/1200/High</td>
<td>12</td>
<td>72</td>
</tr>
<tr>
<td>8</td>
<td>8/900/High</td>
<td>16</td>
<td>96</td>
</tr>
<tr>
<td>6</td>
<td>4/1500/High</td>
<td>20</td>
<td>120</td>
</tr>
<tr>
<td>10</td>
<td>8/1200/High</td>
<td>23</td>
<td>138</td>
</tr>
<tr>
<td>14</td>
<td>12/900/High</td>
<td>25</td>
<td>150</td>
</tr>
<tr>
<td>16</td>
<td>12/1200/High</td>
<td>36</td>
<td>216</td>
</tr>
<tr>
<td>12</td>
<td>8/1500/High</td>
<td>40</td>
<td>240</td>
</tr>
<tr>
<td>18</td>
<td>12/1500/High</td>
<td>60</td>
<td>360</td>
</tr>
<tr>
<td>1</td>
<td>4/900/Low</td>
<td>12</td>
<td>72</td>
</tr>
<tr>
<td>3</td>
<td>4/1200/Low</td>
<td>19</td>
<td>114</td>
</tr>
<tr>
<td>5</td>
<td>4/1500/Low</td>
<td>20</td>
<td>120</td>
</tr>
<tr>
<td>7</td>
<td>8/900/Low</td>
<td>24</td>
<td>144</td>
</tr>
<tr>
<td>9</td>
<td>8/1200/Low</td>
<td>34</td>
<td>204</td>
</tr>
<tr>
<td>13</td>
<td>12/900/Low</td>
<td>35</td>
<td>210</td>
</tr>
<tr>
<td>11</td>
<td>8/1500/Low</td>
<td>40</td>
<td>240</td>
</tr>
<tr>
<td>15</td>
<td>12/1200/Low</td>
<td>46</td>
<td>276</td>
</tr>
<tr>
<td>17</td>
<td>12/1500/Low</td>
<td>60</td>
<td>360</td>
</tr>
</tbody>
</table>

Note. The rows labeled “Kills” and “DMPIs” show the maximum possible number of each variable for that particular scenario. Note that the maximum number of DMPIs is 6 times the maximum number of kills. The data points shown in Figures 16-19 match the scenario sequences listed above (as compared to the sequence shown in Table 1).

High Attack Difficulty. Figure 16 shows the mean number of kills is plotted against the maximum number of kills for the high attack difficulty scenarios. The nine data points were obtained from the nine conditions listed in the top of Table 4. For example, the first three data points represent scenarios #2, #4 and #8, respectively. In scenario #8, the operators were supervising eight UCAVs flying at 900 knots. The mean number of kills for scenario #8 was 11.33 compared to a maximum of 16. Accomplishment curves have been adequately fit by quadratic functions (Colle & Reid, 2005) and by piecewise linear functions (Colle & Reid, 1998, 2006). Therefore, the best-fitting function was sought empirically. The maximum of this function can then be used algebraically to estimate the peak performance value.
The high attack difficulty data were analyzed using a repeated-measures ANOVA with orthogonal polynomial contrasts with uneven spacing to determine which functions might best fit the data. The linear, quadratic, and cubic functions were all statistically significant, $F(1, 11) = 18.71, \text{MSE} = 9.59, p = .001, F(1, 11) = 41.32, \text{MSE} = 3.96, p < .0001$, and $F(1, 11) = 19.95, \text{MSE} = 2.79, p = .001$, respectively. The complete ANOVA table is shown in Appendix P.

Because quadratic functions did not adequately describe the data, and because the data in Figure 16 appear to flatten as they had in Colle and Reid (1998, 2006), a piecewise linear function was used. This is a special case of a spline and knot function. The first spline is linear with an intercept of zero. The spline after the knot is flat (slope of zero, intercept equal to the maximum). The function that was estimated was

$$K = \begin{cases} 
  b \cdot P & \text{if } P < M \\
  c & \text{if } P \geq M 
\end{cases}$$

Eq. (1)

where $K =$ estimated mean number of kills, $P =$ maximum possible number of kills, $b =$ slope of the first spline, and $c =$ the mean of the second spline and the y-value of the knot. $M$ represents the maximum number of kills beyond which operator performance does not improve, (i.e., the performance redline). It is the x-value of the knot. The parameters of $b$ and $c$ were estimated simultaneously using a least squares procedure. The least squares estimator of the slope of the linear spline function with intercept zero is $\Sigma K \cdot P / \Sigma P^2$ for the subset of points before the knot. The least squares estimator for the flat spline function after the knot is the mean of the subset of points after the knot. Thus, the procedure estimated the slope for 1, 2, 3, etc. points from the left and the mean of the remaining points. The sum of the squared deviation of all 9 observed data points from the points predicted by the estimated $b$ and $c$ parameters was obtained, $\Sigma (Y - K)^2$, 

50
where $Y$ is mean number of kills and $K$ is the estimated mean number of kills. Ten estimates of the sum of the squared deviations from the predicted function were obtained by placing the knot in all ten possible locations between points or at the high end. The best-fitting estimates of $b$ and $c$ were the values at the knot location, which produced the minimum of the sum of squared deviations. The sum of squared deviations from the predicted function for all 10 knot positions is provided in Appendix Q, along with the corresponding estimates of $b_i$ and $c_i$, where $i$ refers to the knot position.

The solid lines in Figure 16 shows the best-fitting piecewise linear function that was obtained. The dashed diagonal line shows the maximum possible performance. The sum of the squared deviation of all 9 observed data points from the points predicted by the estimated $b$ and $c$ parameters was obtained. The minimum sum of squared deviations from the predicted function was $\sum(Y-K)^2 = 1.9431$, where $Y$ is mean number of kills and $K$ is the estimated mean number of kills. The $R^2$ for the relationship between $Y$ and $K$ was 0.9479. The slope of the first spline, $b$, was .73 and mean of the second spline, $c$, was 11.90. The performance redline, $M$, is the $x$-value of $P$ that produces $c$ in Eq. (1). It was found by algebraic computation to be 16.34. The arrow in Figure 16 shows this point graphically. This knot point is near scenario #8. Thus, with the high difficulty task, maximal performance was obtained with 8 aircraft flying at 900 knots. Increases in airspeed and number of UCAVs supervised beyond that did not increase the number of kills accomplished, although points nearby may also be considered, depending on mission objectives. The discussion section considers these issues more fully.
The analysis of the DMPIs in the high attack difficulty condition followed the same procedure as the kills. Because the number of DMPIs followed the number of kills, similar results were obtained. Polynomial ANOVAs on the DMPIs revealed statistically significant results for the first through fourth order functions, $F(1, 11) = 49.81, MSE = 375.98, p = < .0001$, $F(1, 11) = 41.51, MSE = 177.65, p = < .0001$, $F(1, 11) = 20.25, MSE = 67.76, p = .001$, and $F(1, 11) = 10.46, MSE = 148.37, p = .007$, respectively. The complete polynomial ANOVA table is listed in Appendix R. A piecewise linear function was again used to fit the data. As shown in Figure 17, the knot of the piecewise linear function again occurs near the third data point. The minimum sum of squared deviations from the predicted function was $\Sigma(Y-K)^2 = 297.542$, where $Y$ is mean number of kills and $K$ is the estimated mean number of kills. The $R^2$ for the relationship between $Y$ and $K$ was 0.9479. The slope of the first spline, $b$, was .73 and mean of the second spline, $c$, was 11.90. The performance redline, $M$, is the x-value of $P$ that produces $c$ in Eq. (1). The performance redline value for DMPIs is 105.52. As indicated in Table 4, the
third data point represents scenario # 8, the high difficulty scenario whereby the operators were supervising 8 UCAVs flying at 900 knots. The fourth data point represents scenario # 6, the high difficulty scenario whereby the operators were supervising 4 UCAVs flying at 1500 knots. The mean number of DMPIs for scenarios 8 and 6 were 80.31 and 86.77, respectively.

![Graph](image)

Figure 17. Accomplishment DMPI score for High Attack Difficulty Scenarios.

**Low Attack Difficulty.** The analysis of the measures used in the low attack difficulty condition followed the same procedure used for the high attack difficulty measures. Similar results were obtained between the number of kills and DMPIs. Polynomial ANOVAs on the kills revealed statistically significant results for the first and second order function, $F(1, 11) = 64.65, \text{MSE} = 74.41, p = < .001$, and $F(1, 11) = 125.31, \text{MSE} = 7.34, p = < .001$. The complete polynomial ANOVA table is listed in Appendix S. A piecewise linear function was again used to fit the data. As shown in Figure 18, the knot of the best-fitting piecewise linear function occurs between the sixth and seventh data points. The performance redline value for kills is 36.66. As indicated in Table 4, the sixth data point represents scenario # 13, the low difficulty
scenario whereby the operators were supervising 12 UCAVs flying at 900 knots. The seventh data point represents scenario # 11, the low difficulty scenario whereby the operators were supervising 8 UCAVs flying at 1500 knots. The mean number of kills for scenarios 13 and 11 were 28.42 and 32.42, respectively.

Polynomial ANOVAs on the DMPIs revealed statistically significant results for the first and second order function, \( F(1, 11) = 136.02, \text{MSE} = 1933.87, p = < .001 \), and \( F(1, 11) = 87.18, \text{MSE} = 379.10, p = < .001 \). The complete polynomial ANOVA table is listed in Appendix T. Once again, a piecewise linear function was used to fit the data. As Figure 19 shows, the knot of the piecewise linear function occurs between the sixth and seventh data points, similar to the kills. The performance redline value for DMPIs is 232.85. As indicated in Table 4, the sixth data point represents scenario # 13, the low difficulty scenario whereby the operators were supervising 12 UCAVs flying at 900 knots. The seventh data point represents scenario # 11, the
low difficulty scenario whereby the operators were supervising 8 UCAVs flying at 1500 knots. The mean number of DMPIs for scenarios 13 and 11 were 189.58 and 207.17, respectively.

![Low Difficulty Accomplishment](image)

**Figure 19.** Accomplishment DMPI scores for Low Attack Difficulty Scenarios.

d. Estimation of Subjective Mental Workload Redlines

*High Difficulty -- SWAT.* The performance redline determined from the high difficulty kills was used to estimate the subjective mental workload redlines. Polynomial ANOVAs on the kills revealed statistically significant results for the first, second, and fourth order functions, \(F(1, 11) = 329.02, MSE = 160.35, p = < .001\), \(F(1, 11) = 12.63, MSE = 284.80, p = .005\), \(F(1, 11) = 8.77, MSE = 118.75, p = .01\), respectively. The complete polynomial ANOVA table is listed in Appendix U. Figure 20 plots the best-fitting linear and polynomial equations. The algebraic best-fitting equations and corresponding \(R^2\) values also are shown in Figure 20. The best-fitting (fourth order) polynomial in the figure contains only the orthogonal polynomial effects that were statistically significant in the polynomial ANOVA. The orthogonal effects of the non-significant third order function have been removed from the best-fitting polynomial equation. The
polynomial equation includes a non-zero coefficient for the third order term because the fourth order effect produced these weights when the orthogonal basis was converted back to the units on the x axis. The $R^2$ for the polynomial equals the sum of the $r^2$ for the significant orthogonal polynomial effects, in this case for orders 1, 2, and 4. Figure 20 also shows the performance redline of 16.34 as a vertical line and the points at which it intercepts the linear and polynomial functions for the high difficulty SWAT ratings. The horizontal lines show the estimates of the SWAT redline. The estimated SWAT redline value is 39.66 for the linear function and 35.47 for the polynomial function when calculated algebraically from these best-fitting functions. These values fall into the redline range of $40 \pm 10$ as previously indicated by Colle and Reid (1988, 2005).

![Graph showing SWAT redlines as calculated from high attack difficulty kills Performance Redline](image)

**Low Difficulty -- SWAT.** The performance redline derived from the low difficulty kills was used to estimate a subjective mental workload redline. Polynomial ANOVAs on the kills
revealed statistically significant results for the first, third, and fifth order functions, $F(1, 11) = 189.58$, $MSE = 228.58$, $p = < .001$, $F(1, 11) = 10.28$, $MSE = 110.78$, $p = .008$, $F(1, 11) = 20.78$, $MSE = 73.84$, $p = < .001$, respectively. The complete polynomial ANOVA table is listed in Appendix V. Figure 21 plots the best-fitting linear and polynomial equations. The best-fitting polynomial in the figure contains only the significant orthogonal polynomial effects obtained in the polynomial ANOVA. The orthogonal effects of the non-significant second and fourth order functions have been removed from the computation. Figure 21 shows the best-fitting equations and corresponding $R^2$ values. Figure 21 also shows a vertical line rising from the performance redline of 36.66. This line intercepts the linear and fifth order polynomial functions to indicate the low difficulty SWAT ratings. The horizontal lines show the SWAT redline estimates. The estimated SWAT redline value is 33.11 for the linear function and 37.04 for the polynomial function when calculated algebraically from these best-fitting functions. These values fall in the redline range of $40 \pm 10$ as previously indicated by Colle and Reid (1988, 2005).

![Figure 21. SWAT redlines as indicated from low attack difficulty kills Performance Redline.](image-url)
High Difficulty – TLX-M4. The performance redline obtained from the high difficulty kills was used to estimate subjective mental workload redlines for the TLX-M4 as well. The high difficulty TLX-M4 workload ratings were analyzed and plotted to establish a TLX-M4 redline rating. Similar to the SWAT results, the polynomial ANOVAs on the kills revealed statistically significant results for the first, second, and fourth order functions on the TLX-M4 ratings, $F(1, 11) = 137.81, MSE = 184.22, p = < .001$, $F(1, 11) = 26.68, MSE = 116.16, p = < .001$, $F(1, 11) = 9.28, MSE = 63.02, p = .01$, respectively. The complete polynomial ANOVA table is listed in Appendix W. As can be seen in Figure 22, the first and fourth order functions for the high difficulty TLX-M4 ratings are graphed with lines intersecting with the performance redline of 16.34. The best-fitting (fourth order) polynomial in the figure contains only the significant effects. The effects of the non-significant cubic function have been removed from the computation. The TLX-M4 redline values calculated from the kills performance redline are 53.77 for the linear function and 50.57 for the polynomial function.

Figure 22. TLX-M4 redlines as indicated from high attack difficulty kills Performance Redline.
Low Difficulty – TLX-M4. The performance redline obtained from the low difficulty kills was used to estimate subjective mental workload redlines for the TLX-M4 ratings as well. The polynomial ANOVAs on the kills revealed statistically significant results for the first, third, fifth, and seventh order functions on the low TLX-M4 ratings, $F(1, 11) = 215.87, MSE = 214.37, p = < .001$, $F(1, 11) = 10.51, MSE = 120.31, p = 0.008$, $F(1, 11) = 14.10, MSE = 45.52, p = .003$, and $F(1, 11) = 20.43, MSE = 74.87, p = < .001$ respectively. The complete polynomial ANOVA table is listed in Appendix Y. Appendices Y and Z show the high difficulty and low difficulty, respectively. As can be seen in Figure 23, the linear and polynomial functions for the low difficulty TLX-M4 ratings were graphed with lines intersecting with the performance redline of 36.66. The best-fitting (seventh order) polynomial in the figure contains only the significant effects obtained from the polynomial ANOVA. The effects of the non-significant quadratic, quartic, and sextic functions have been removed from the computation. The percent of variance explained by the function is high (98.33%). It should be noted that the polynomial function does not appear to be conceptually reasonable. The polynomial function goes into a negative range twice, and also rises above 100 to over 1,000 between the 8th and final data point. The TLX-M4 redline values calculated from the kills performance redline are 48.77 for the linear function and 63.02 for the polynomial function.
V. Discussion

a. Summary of Research Goals

As outlined in the introduction (Section III), this study had four major goals. The first goal was to determine the effects of number of UCAVS, vehicle airspeed, and difficulty level of attacks (targets engaged by either a single vehicle or multiple vehicles) on performance during a SEAD mission scenario. The operators were trained to discriminate between target vehicles and distractors, while destroying as many target icons as possible. A minimal number of distractors were destroyed, only 0.04% for all participants across all trials, indicating that the operators were sufficiently trained to discriminate between the two vehicle types. Given that very few distractors were destroyed, the performance data analyses focused on standard performance measures, the percentage of kills and correctly-placed DMPIs. The three independent variables were combined factorially and all three factors were found to have large effects on the percentage of kills and DMPIs.
However, these factorial analyses do not provide a simple answer to the question of how many simulated UCAVs one operator may effectively manage at different airspeeds and task difficulty levels. This second goal was addressed using Colle and Reid’s approach of defining accomplishment scores for a fixed time interval and then estimating performance redlines. For high difficulty levels, the redline occurred near scenario #8, the high difficulty scenario whereby the operators were supervising 8 UCAVs flying at 900 knots. Scenario #6 had similar accomplishment performance to scenario #8’s, in which operators supervised 4 UCAVs flying at 1500 knots. Note that although the maximum possible number of kills and DMPIs increased considerably above these accomplishment score levels and the percentage of kills and DMPIs also decreased considerably, the accomplishment scores were flat after the redline. Accomplishment score analysis demonstrated results not plainly shown through factorial analysis.

In the low difficulty scenarios, the redline occurred near scenario #13, in which operators were supervising 12 UCAVs flying at 900 knots. The following scenario, #11, involved the operators supervising 8 UCAVs flying at 1500 knots. Other nearby points represent parameters that may be considered during the formation of mission objectives. Here the results of the accomplishment score analysis more obviously related to the factorial analysis. For example, in Figures 9 and 11, Kill % and Correctly-Placed DMPIs % for Low Attack Difficulty, respectively, there is a ceiling effect. The operators had near-perfect performance while supervising four UCAVs. Overall, however, a distinctive point of performance breakdown can be difficult to determine from factorial ANOVAs. In the two figures showing High Attack Difficulty results for Kills and Correctly-placed DMPIs (Figures 8 and 10, respectively), the uncertainty of a point of performance breakdown is exacerbated by the overlapping scores along the y-axis.
Conversely, when one interprets Figures 16-19, Accomplishment Kill and DMPI Scores for High and Low Attack Difficulty, respectively, one is quickly drawn to the distinctive thresholds created by the performance redline. The divergence of the solid lines, (indicating the best-fitting piecewise linear function) from the dashed lines (indicating the maximum possible performance), along with the “knee”, provide a clearer indication of apical performance. An estimation of maximal operator performance in this domain, given the parameters on and around the performance redline, is more straightforward.

A third goal was to test the previously-established SWAT redline estimate of 40 ±10 in a simulated UCAV domain. A SWAT redline can be equated to a car tachometer that indicates when an engine is operating at a dangerously high level. Continued operation at or above this redline might lead to performance breakdown. The performance redlines that were estimated for each difficulty level were related to SWAT ratings for the low and high difficulty levels when plotted against the maximal number of possible kills. These redlines were then used to estimate mental workload redlines. For the high difficulty level the SWAT redline estimates were 39.66 and 35.47 for the linear and polynomial estimates, respectively. The linear estimate accounted for most of the variance and was consistent with the linear estimates used by Colle and Reid (2005), who did not find higher order effects with the polynomial ANOVA. Both the linear and the polynomial effects are consistent with the 40 ±10 redline rule proposed by Reid and Colle (1988). A similar analysis of low difficulty levels found SWAT redline estimates of 33.11 and 37.04 for the linear and polynomial estimates, respectively. Both of these estimates again were within the 40 ±10 redline rule. This estimate is important because it provides support for global redline invariance, which Reid and Colle (2005) emphasized was necessary for a redline to achieve maximum usability. Note that the high and low difficulty levels were obtained with very
different performance redlines and performance generally. Thus, the present results not only
provided support for the 40 ±10 redline rule in a new domain, the control of multiple UCAVs,
they also showed within the same experimental approach and study that equivalent SWAT
redlines were estimated for very different difficulty conditions. These SWAT redline estimates
and performance redlines for the low and high attack difficulty levels provide converging
support for identifying the performance threshold in this simulation environment.

A fourth and final goal of the present study was to provide the first redline estimate for
the NASA TLX. Similar to the SWAT redline procedure, the performance redlines that were
estimated for each difficulty level were related to TLX-M4 ratings for the low and high difficulty
levels when plotted against the maximal number of possible kills. These redlines were used to
estimate mental workload redline estimates for the TLX-M4. For the high difficulty level, the
TLX-M4 redline estimates were 53.77 for the linear function and 50.57 for the polynomial
function. Analyses of the low difficulty levels found TLX-M4 redline estimates of 48.77 and
63.02 for the linear and polynomial functions, respectively. I propose that a TLX-M4 redline
could be considered in the range of 50 ± 10 for this simulation environment. The estimates that
comprise this range are developed from a technique similar to that used in estimating the SWAT
redline. This redline estimate excludes the estimate from the low attack difficulty polynomial
function (63.02) shown in Figure 23, which was an atypical function because it violated expected
boundary conditions of the TLX-M4 scale. The present analysis focused on the M4 version of
the TLX for two reasons. First, the physical workload and performance subscales were excluded
because they do not have face validity as subjective mental workload scales. Second, the other
four subscales were combined so that the overall weighted scale would be more likely to possess
global redline invariance. Some evidence for this was found given that the linear redline TLX-
M4 estimates were 54 versus 49 for the high and low attack difficulty conditions. However, there may be differences on the individual subscales as can be seen in Appendix AC, especially for low attack difficulty scenarios. Unlike SWAT, which produces one combined scale, individual subscales of TLX have been used to interpret mental workload differences between experimental conditions. Therefore, the viability of TLX-M4 and the individual subscale redlines needs to be explored further.

b. Limitations and Future Research

There were several limitations on the present study that should be considered during future research. Ten of the twelve operators had experience on UCAV or UAV simulators. However, even the more experienced operators were unlikely able to achieve the same degree of realism as an actual UCAV operator engaged in a real mission. This factor exists in most, if not all, training and simulation environments. The operational mission environment and the commensurate level of stress may affect operator performance differently.

Another limitation involves the definitions of high and low task difficulty. The high difficulty level required the operators to engage each target icon with at least two UCAVs, while the low difficulty level required the operators to use only one UCAV per target icon. In addition, the high difficulty level had non-target distracters while the low difficulty level did not. Although the vehicle speed and number of vehicles quantitatively spanned the range of variation that was likely to be considered in operational settings, the difficulty variable had only the above two qualitatively different levels. The present study used variables that allowed the operators to perform tasks that have been identified as those most likely to be employed for an operational system. When investigating performance thresholds for UCAV operators, many manipulations could be used to distinguish the difficulty levels of a primary task. While vehicle airspeed and
the number of vehicles to be supervised are likely to remain important factors, so will identifying the operator’s primary task responsibilities.

Along those lines, it’s also interesting to consider the different configurations of operation. For example, it has also not yet been determined if one operator will manage aircraft health and maintenance tasks along with the targeting duties. Not only are there different sub-tasks that an operator could perform, there are also different roles an operator could perform. For example, remember that two basic roles exist in controlling UCAVs: a pilot/operator and a sensor operator. Given that these two roles could conceivably be performed by one or more operators, a question arises as to the optimal number of roles and operators per aircraft. Will one operator be able to manage the multiple sub-tasks involved, such as piloting/supervising, sensor operation, targeting, health status alerts, etc., in an un-aided system? Will one operator or supervisor manage the health status alerts for several operators? Will a portion of the system become adaptively automated at times of performance breakdown to help the operator recover? How would that system be invoked (e.g., performance measures, psychophysiological triggers, contextual cues, operator-activated, or some combination of triggers)? How much automation is optimal?

Sheridan (2006) discussed several variations of the term Human-Centered Automation. One of the meanings he listed has particular relevance to this thesis: Allocate to the human the task(s) best suited to the human (i.e., give final clearance for the release of munitions) and allocate to the system the task(s) best suited to it (i.e., flying from waypoint to waypoint). This concept is the underlying motivation for the present study, given that the primary task demands of UCAV operators will likely center around placing bombs on targets as accurately and efficiently as possible.
Future UCAV workload studies should first seek to further explore the principal factors that contribute to performance breakdown in operational scenarios. As the UCAV roles and responsibilities become more defined, future studies should further explore the task difficulty as well. The two levels of task difficulty in the present study were largely defined by the number of vehicles used to engage a target icon. Inherent to these conditions was a subset of cognitive tasks including visual search, situational awareness, context switching, as well as the physical movements used to manipulate the input device. Identifying and mitigating these sources of workload in the UCAV domain requires more specific classification of both the primary and secondary tasks that may be involved.

c. Potential Applications

Clearly, no single study is able to address all of these limitations. However, subsequent studies can build upon the lessons learned from this study and other previous research. For example, in addition to the UCAV program, another well-documented program which sought to reduce mental workload, human error, and the number of pilots necessary to carry out a mission was the Pilot’s Associate (PA) program, which was active in the 1980’s and 1990’s. The Air Force PA program, which DARPA began in 1986, used a series of algorithms, sensors and pre-determined “contracts” to determine when a task would be temporarily managed by an on-board computer. This computer sought to generate an additional “associate” in the single-seat fighter plane cockpit, which could help to manage threat assessment, deal with threat engagement and avoidance, inform about aircraft health status, and conduct dynamic mission re-planning (Banks & Lizza, 1991).

The PA system consisted of two planners, two assessors and a pilot-vehicle interface. The tactics planner and the mission planner responded to the events in the dynamic environment
as they affected the pre-briefed mission plan. The situation assessor and the system status assessor helped to determine the state of the outside world and the aircraft, respectively. Just as in the case of the UCAV, the user interface was the key link between the system and the operator. The interface was responsible for managing the communication between the pilot and the various modules. The system would check for critical errors and assume control of certain tasks when the pilot was overloaded, if the task re-allocation had been authorized prior to the mission (Banks & Lizza, 1991).

Svenmarck and Dekker (2003) noted that the pilots appreciated the adaptive system and the way it presented information, as well as how it provided integration of sensor inputs. During most simulated missions, the pilots did not need to override the choice or timing of the displays (Banks & Lizza 1991). However, there were some complaints with the way the system occasionally misinterpreted the intentions of the pilot. Also, in trying to provide the optimal distribution of tasks, the system would sometimes conflict with the pilot’s actions and strategies. By mismanaging the task allocation in a time-critical and dynamic battlefield environment, the PA would at times create assignment shifts that increased the pilot’s mental workload (Svenmarck and Dekker 2003). However, subsequent improvements to the algorithms that managed the shifts reduced these effects on the pilot’s performance (J.M. Reising, personal communication, October 5, 2005).

This platform later evolved into the Rotorcraft Pilot’s Associate (RPA) program, tested by the U.S. Army in their helicopters, which in turn, led to DARPA’s AugCog program. The AugCog program (also referred to as Improving Warfighter Information Intake Under Stress) seeks to develop a closed-loop interface that adapts the system to the user’s needs during periods of stress and/or degraded performance. By identifying the factors that most contribute to
performance degradation, the lessons learned from efforts such as PA, RPA and UAV programs may culminate in an optimal user interface for the U.S. military’s UCAV program.
References


Illinois, Aviation Research Laboratory.


Appendix A

NASA TLX dimensions

1. Mental Demand: How much mental and perceptual activity was required? Was the task easy or difficult, simple or complex, rigorous or forgiving?

2. Physical Demand: How much physical activity was required? Was the task easy or difficult, slow or fast-paced, restful or laborious?

3. Temporal Demand: How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

4. Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?

5. Performance: How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance?

6. Frustration: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, relaxed and complacent did you feel during the task?
Appendix B

Subjective Workload Assessment Technique (SWAT) Dimensions

I. Time Load

1. The operator often has spare time. Interruptions or overlap among activities occur infrequently or not at all.

2. The operator occasionally has spare time. Interruptions or overlap among activities occur frequently.

3. The operator almost never has spare time. Interruptions or overlap among activities are very frequent, or occur all the time.

II. Mental Effort Load

1. There is very little conscious mental effort or concentration required. Activity is almost automatic, requiring little or no attention.

2. There is moderate conscious mental effort or concentration required. The complexity of mental activity is moderately high due to uncertainty, unpredictability, or unfamiliarity. Considerable attention is required.

3. Extensive mental effort and concentration are required. Very complex activity requiring total attention.

III. Psychological Stress Load

1. There is little confusion, risk, frustration, or anxiety, and is easily accommodated.

2. The moderate stress due to confusion, frustration, or anxiety noticeably adds to workload. Significant compensation is required to maintain adequate performance.

3. High to very intense stress is present due to confusion, frustration, or anxiety.
## Appendix C

### Mean TLX-M4 Rating scale

<table>
<thead>
<tr>
<th>Scenario #</th>
<th>Mean Composite TLX-M4 Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13.82</td>
</tr>
<tr>
<td>2</td>
<td>36.18</td>
</tr>
<tr>
<td>3</td>
<td>17.50</td>
</tr>
<tr>
<td>4</td>
<td>45.56</td>
</tr>
<tr>
<td>5</td>
<td>29.44</td>
</tr>
<tr>
<td>6</td>
<td>58.40</td>
</tr>
<tr>
<td>7</td>
<td>22.78</td>
</tr>
<tr>
<td>8</td>
<td>52.08</td>
</tr>
<tr>
<td>9</td>
<td>38.96</td>
</tr>
<tr>
<td>10</td>
<td>63.75</td>
</tr>
<tr>
<td>11</td>
<td>62.57</td>
</tr>
<tr>
<td>12</td>
<td>74.58</td>
</tr>
<tr>
<td>13</td>
<td>53.89</td>
</tr>
<tr>
<td>14</td>
<td>74.38</td>
</tr>
<tr>
<td>15</td>
<td>66.67</td>
</tr>
<tr>
<td>16</td>
<td>81.88</td>
</tr>
<tr>
<td>17</td>
<td>75.35</td>
</tr>
<tr>
<td>18</td>
<td>90.07</td>
</tr>
</tbody>
</table>

### TLX-M4 Sources of Workload weights

<table>
<thead>
<tr>
<th>Subject</th>
<th>Mental Demand</th>
<th>Temporal Demand</th>
<th>Effort</th>
<th>Frustration</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>H</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>I</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>J</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>K</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>L</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Mean</td>
<td>1.58</td>
<td>1.83</td>
<td>1.58</td>
<td>1.17</td>
</tr>
</tbody>
</table>

Note: Divide mean by 6 to obtain proportional weights
### Mean TLX Rating scale – All 6 factors

<table>
<thead>
<tr>
<th>Scenario #</th>
<th>Mean Composite TLX Rating – All 6 factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.00</td>
</tr>
<tr>
<td>2</td>
<td>32.53</td>
</tr>
<tr>
<td>3</td>
<td>15.22</td>
</tr>
<tr>
<td>4</td>
<td>43.97</td>
</tr>
<tr>
<td>5</td>
<td>24.61</td>
</tr>
<tr>
<td>6</td>
<td>55.31</td>
</tr>
<tr>
<td>7</td>
<td>19.44</td>
</tr>
<tr>
<td>8</td>
<td>48.61</td>
</tr>
<tr>
<td>9</td>
<td>34.75</td>
</tr>
<tr>
<td>10</td>
<td>62.44</td>
</tr>
<tr>
<td>11</td>
<td>57.97</td>
</tr>
<tr>
<td>12</td>
<td>72.64</td>
</tr>
<tr>
<td>13</td>
<td>46.86</td>
</tr>
<tr>
<td>14</td>
<td>70.42</td>
</tr>
<tr>
<td>15</td>
<td>60.39</td>
</tr>
<tr>
<td>16</td>
<td>77.17</td>
</tr>
<tr>
<td>17</td>
<td>69.22</td>
</tr>
<tr>
<td>18</td>
<td>87.14</td>
</tr>
</tbody>
</table>
Appendix D

Training Procedure

1. The proposed role of the UCAV and its SEAD mission was described. The SAR targets and distracter objects were also described.

2. The TSD was shown to the operators and a broad overview of the tasks (i.e., SAR-target icon pairings, DMPI placement, munition assignments, etc.) was presented.

3. The operator practiced using four UCAVs flying at 800 knots with only 12 target icons available on the TSD. The SAR images contained only C-targets. I then flew a scenario and described the strategies that would be helpful to their performance, and often paused the task to describe the current situation to the participants.

4. As the operators progressed through this mission level (and throughout the training process), the rules of engagement were outlined as in the Procedure and Task sections.

5. After the operators mastered this easy level without error, they were introduced to 8-UCAV and 12-UCAV scenarios with easy SAR images, flying at 800 knots.

6. Afterwards, the operators were introduced to 4-, 8-, 12-UCAV high attack difficulty scenarios, flying at 1300 knots.

7. After the operators were introduced to these scenarios, they were introduced to three Final Training scenarios, using the two endpoint parameters (i.e., 4 UCAVs, 900 knots, low attack difficulty, and 12 UCAVs, 1500 knots, high difficulty, respectively), and a middle level of difficulty (i.e., 8 UCAVs, 1200 knots, high difficulty).

8. During step 7, the operators filled out the TLX forms and SWAT forms, in a counterbalanced order, representative of their Operator Number from Table 2.

9. After the final scenario, the operators completed the TLX Sources of Workload survey.
# Appendix E

## Numeric SWAT scale estimated from the card sort and conjoint analysis

<table>
<thead>
<tr>
<th>Card label</th>
<th>Rating Combination</th>
<th>Rescaled value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1 1 1 0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>B</td>
<td>1 1 2 15.4</td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>1 1 3 34.8</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>1 2 1 15.4</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>1 2 2 30.7</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1 2 3 50.1</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>1 3 1 34.3</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>1 3 2 49.6</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>1 3 3 69.0</td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>2 1 1 17.1</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>2 1 2 32.5</td>
<td></td>
</tr>
<tr>
<td>Z</td>
<td>2 1 3 51.9</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>2 2 1 32.5</td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>2 2 2 47.8</td>
<td></td>
</tr>
<tr>
<td>ZZ</td>
<td>2 2 3 67.2</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>2 3 1 51.4</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>2 3 2 66.7</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>2 3 3 86.1</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>3 1 1 31.0</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>3 1 2 46.4</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>3 1 3 65.7</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>3 2 1 46.3</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>3 2 2 61.7</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>3 2 3 81.1</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>3 3 1 65.2</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>3 3 2 80.6</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>3 3 3 100.0</td>
<td></td>
</tr>
</tbody>
</table>
Appendix F

Summary of distracters destroyed in each high attack difficulty scenario

<table>
<thead>
<tr>
<th>Scenario #</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Number of distracters present

<table>
<thead>
<tr>
<th>Scenario #</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>15</th>
<th>3</th>
<th>3</th>
<th>0</th>
<th>0</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>387</td>
<td>387</td>
<td>387</td>
<td>774</td>
<td>771</td>
<td>774</td>
<td>1170</td>
<td>1163</td>
<td>1161</td>
</tr>
<tr>
<td>2</td>
<td>387</td>
<td>387</td>
<td>387</td>
<td>774</td>
<td>771</td>
<td>774</td>
<td>1170</td>
<td>1163</td>
<td>6974</td>
</tr>
</tbody>
</table>

Note. The total number of distracters present for all 9 scenarios was 83,688. There were no distracters in the low attack difficulty scenarios.
## Appendix G

### Percent Kills ANOVA table for high and low difficulty

<table>
<thead>
<tr>
<th>Source</th>
<th>SSQ</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\epsilon_{F-G}$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (S)</td>
<td>132734.44</td>
<td>11</td>
<td>12066.77</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Difficulty (D)</td>
<td>66969.18</td>
<td>1</td>
<td>66969.18</td>
<td>404.6</td>
<td>---</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x D</td>
<td>1820.7</td>
<td>11</td>
<td>165.52</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Number of vehicles (V)</td>
<td>38488.76</td>
<td>2</td>
<td>19244.38</td>
<td>94.79</td>
<td>.74</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x V</td>
<td>4466.64</td>
<td>22</td>
<td>203.03</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Airspeed (A)</td>
<td>20303.52</td>
<td>2</td>
<td>10151.76</td>
<td>407.87</td>
<td>.88</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x A</td>
<td>547.49</td>
<td>22</td>
<td>24.89</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>V x D</td>
<td>1825.64</td>
<td>2</td>
<td>912.82</td>
<td>8.45</td>
<td>.86</td>
<td>= .003</td>
</tr>
<tr>
<td>S x V x D</td>
<td>2376.36</td>
<td>22</td>
<td>108.02</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A x D</td>
<td>1850.01</td>
<td>2</td>
<td>925.0</td>
<td>10.11</td>
<td>.85</td>
<td>= .002</td>
</tr>
<tr>
<td>S x A x D</td>
<td>2011.95</td>
<td>22</td>
<td>91.45</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>V x A</td>
<td>1667.57</td>
<td>4</td>
<td>416.89</td>
<td>8.78</td>
<td>.73</td>
<td>= .003</td>
</tr>
<tr>
<td>S x V x A</td>
<td>2089.99</td>
<td>44</td>
<td>47.5</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A x V x D</td>
<td>1629.76</td>
<td>4</td>
<td>407.44</td>
<td>8.04</td>
<td>.60</td>
<td>= .0002</td>
</tr>
<tr>
<td>S x A x V x D</td>
<td>2230.27</td>
<td>44</td>
<td>50.69</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>
**Appendix H**

**Percent Kills ANOVA table for high difficulty**

<table>
<thead>
<tr>
<th>Source</th>
<th>SSQ</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\bar{\epsilon}_{e.g.}$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (S)</td>
<td>43027.87</td>
<td>11</td>
<td>3911.625</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Number of vehicles (V)</td>
<td>25402.83</td>
<td>2</td>
<td>12701.41</td>
<td>115.45</td>
<td>.92</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x V</td>
<td>2420.32</td>
<td>22</td>
<td>110.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Airspeed (A)</td>
<td>17000.48</td>
<td>2</td>
<td>8500.24</td>
<td>106.55</td>
<td>.91</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x A</td>
<td>1755.08</td>
<td>22</td>
<td>79.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x V</td>
<td>624.56</td>
<td>4</td>
<td>156.14</td>
<td>2.71</td>
<td>.65</td>
<td>= .07</td>
</tr>
<tr>
<td>S x A x V</td>
<td>2539.33</td>
<td>44</td>
<td>57.71</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix I

Percent Kills ANOVA table for low difficulty

<table>
<thead>
<tr>
<th>Source</th>
<th>SSQ</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>( \epsilon_{GR} )</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (S)</td>
<td>22737.38</td>
<td>11</td>
<td>2067.03</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Number of vehicles (V) S x V</td>
<td>14911.57</td>
<td>2</td>
<td>7455.79</td>
<td>37.09</td>
<td>.62</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Airspeed (A) S x A A x V</td>
<td>5153.05</td>
<td>2</td>
<td>2576.52</td>
<td>70.47</td>
<td>.85</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>A x V</td>
<td>804.36</td>
<td>22</td>
<td>36.56</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S x A x V</td>
<td>1780.94</td>
<td>44</td>
<td>40.48</td>
<td>16.51</td>
<td>.67</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>
## Appendix J

### Percent Correct DMPIs ANOVA table for high and low difficulty

<table>
<thead>
<tr>
<th>Source</th>
<th>SSQ</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>(\varepsilon_{F,d} )</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (S)</td>
<td>109569.54</td>
<td>11</td>
<td>9960.87</td>
<td>---</td>
<td>---</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Difficulty (D)</td>
<td>46906.68</td>
<td>1</td>
<td>46906.68</td>
<td>659.97</td>
<td>---</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x D</td>
<td>781.81</td>
<td>11</td>
<td>71.07</td>
<td>---</td>
<td>---</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Number of vehicles (V)</td>
<td>32564.68</td>
<td>2</td>
<td>16282.34</td>
<td>164.61</td>
<td>.83</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x V</td>
<td>2176.09</td>
<td>22</td>
<td>98.91</td>
<td>---</td>
<td>---</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Airspeed (A)</td>
<td>19798.09</td>
<td>2</td>
<td>9899.05</td>
<td>637.17</td>
<td>.98</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x A</td>
<td>341.79</td>
<td>22</td>
<td>15.54</td>
<td>---</td>
<td>---</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>V x D</td>
<td>3969.88</td>
<td>2</td>
<td>1984.94</td>
<td>34.64</td>
<td>.73</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x V x D</td>
<td>1260.46</td>
<td>22</td>
<td>57.29</td>
<td>---</td>
<td>---</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>A x D</td>
<td>2360.96</td>
<td>2</td>
<td>1180.48</td>
<td>21.42</td>
<td>.87</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x A x D</td>
<td>1212.33</td>
<td>22</td>
<td>55.11</td>
<td>---</td>
<td>---</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>V x A</td>
<td>2202.25</td>
<td>4</td>
<td>550.56</td>
<td>17.24</td>
<td>.72</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x V x A</td>
<td>1405.45</td>
<td>44</td>
<td>31.94</td>
<td>---</td>
<td>---</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>A x V x D</td>
<td>1767.0</td>
<td>4</td>
<td>441.75</td>
<td>14.89</td>
<td>.68</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x A x V x D</td>
<td>1305.27</td>
<td>44</td>
<td>29.67</td>
<td>---</td>
<td>---</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

84
Appendix K

Percent Correct DMPIs ANOVA table for high difficulty

<table>
<thead>
<tr>
<th>Source</th>
<th>SSQ</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\epsilon_{G-G}$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (S)</td>
<td>47300.63</td>
<td>11</td>
<td>4300.06</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Number of vehicles (V)</td>
<td>28711.28</td>
<td>2</td>
<td>14355.64</td>
<td>220.38</td>
<td>.88</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x V</td>
<td>1433.1</td>
<td>22</td>
<td>65.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Airspeed (A)</td>
<td>17670.44</td>
<td>2</td>
<td>8835.22</td>
<td>234.35</td>
<td>.71</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x A</td>
<td>829.44</td>
<td>22</td>
<td>37.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x V</td>
<td>918.91</td>
<td>4</td>
<td>229.73</td>
<td>6.59</td>
<td>.69</td>
<td>= .001</td>
</tr>
<tr>
<td>S x A x V</td>
<td>1533.32</td>
<td>44</td>
<td>34.85</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

85
Appendix L

Percent Correct DMPIs ANOVA table for low difficulty

<table>
<thead>
<tr>
<th>Source</th>
<th>SSQ</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\varepsilon_{p-r}$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (S)</td>
<td>15362.23</td>
<td>11</td>
<td>1396.57</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Number of vehicles (V)</td>
<td>7823.28</td>
<td>2</td>
<td>3911.64</td>
<td>42.95</td>
<td>.62</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x V</td>
<td>2003.45</td>
<td>22</td>
<td>91.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Airspeed (A)</td>
<td>4488.61</td>
<td>2</td>
<td>2244.31</td>
<td>68.13</td>
<td>.84</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x A</td>
<td>724.68</td>
<td>22</td>
<td>32.94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x V</td>
<td>3050.34</td>
<td>4</td>
<td>762.59</td>
<td>28.5</td>
<td>.56</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x A x V</td>
<td>1177.4</td>
<td>44</td>
<td>26.76</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Appendix M

#### SWAT Rating ANOVA table for high and low difficulty

<table>
<thead>
<tr>
<th>Source</th>
<th>SSQ</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>ε_{FG}</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (S)</td>
<td>148125.24</td>
<td>11</td>
<td>13465.93</td>
<td>---</td>
<td>---</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Difficulty (D)</td>
<td>41369.97</td>
<td>1</td>
<td>41369.97</td>
<td>82.86</td>
<td>---</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x D</td>
<td>5492.36</td>
<td>11</td>
<td>499.31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of vehicles (V)</td>
<td>75276.81</td>
<td>2</td>
<td>37638.41</td>
<td>137.37</td>
<td>.91</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x V</td>
<td>6027.79</td>
<td>22</td>
<td>273.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Airspeed (A)</td>
<td>26782.11</td>
<td>2</td>
<td>13391.05</td>
<td>78.63</td>
<td>.78</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x A</td>
<td>3746.8</td>
<td>22</td>
<td>170.31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V x D</td>
<td>890.35</td>
<td>2</td>
<td>445.18</td>
<td>1.3</td>
<td>.88</td>
<td>= .29</td>
</tr>
<tr>
<td>S x V x D</td>
<td>7529.42</td>
<td>22</td>
<td>342.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x D</td>
<td>33.12</td>
<td>2</td>
<td>16.56</td>
<td>.2</td>
<td>.73</td>
<td>= .75</td>
</tr>
<tr>
<td>S x A x D</td>
<td>1805.18</td>
<td>22</td>
<td>82.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V x A</td>
<td>2237.83</td>
<td>4</td>
<td>559.46</td>
<td>2.9</td>
<td>.54</td>
<td>= .07</td>
</tr>
<tr>
<td>S x V x A</td>
<td>8495.43</td>
<td>44</td>
<td>193.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x V x D</td>
<td>1535.05</td>
<td>4</td>
<td>383.76</td>
<td>1.82</td>
<td>.68</td>
<td>= .17</td>
</tr>
<tr>
<td>S x A x V x D</td>
<td>9291.56</td>
<td>44</td>
<td>211.17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix N

TLX Rating All 6 Subscales ANOVA table for High and Low Difficulty

<table>
<thead>
<tr>
<th>Source</th>
<th>SSQ</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\epsilon_{95%}$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (S)</td>
<td>103002.64</td>
<td>11</td>
<td>9363.88</td>
<td>---</td>
<td>---</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Difficulty (D)</td>
<td>29330.04</td>
<td>1</td>
<td>29330.04</td>
<td>114.961</td>
<td>---</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x D</td>
<td>2806.41</td>
<td>11</td>
<td>255.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of vehicles (V)</td>
<td>51784.76</td>
<td>2</td>
<td>25892.38</td>
<td>238.22</td>
<td>.91</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x V</td>
<td>2391.27</td>
<td>22</td>
<td>108.69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Airspeed (A)</td>
<td>18802.78</td>
<td>2</td>
<td>9401.39</td>
<td>105.92</td>
<td>.82</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x A</td>
<td>1952.8</td>
<td>22</td>
<td>88.76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V x D</td>
<td>479.64</td>
<td>2</td>
<td>239.82</td>
<td>2.04</td>
<td>.69</td>
<td>.17</td>
</tr>
<tr>
<td>S x V x D</td>
<td>2591.37</td>
<td>22</td>
<td>117.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x D</td>
<td>132.22</td>
<td>2</td>
<td>66.11</td>
<td>1.31</td>
<td>.81</td>
<td>.29</td>
</tr>
<tr>
<td>S x A x D</td>
<td>1113.31</td>
<td>22</td>
<td>50.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V x A</td>
<td>1333.59</td>
<td>4</td>
<td>333.4</td>
<td>5.86</td>
<td>.69</td>
<td>.003</td>
</tr>
<tr>
<td>S x V x A</td>
<td>2504.06</td>
<td>44</td>
<td>56.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x V x D</td>
<td>1139.61</td>
<td>4</td>
<td>284.9</td>
<td>2.81</td>
<td>.53</td>
<td>.08</td>
</tr>
<tr>
<td>S x A x V x D</td>
<td>4460.45</td>
<td>44</td>
<td>101.37</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix O

TLX-M4 Rating ANOVA table for High and Low difficulty

<table>
<thead>
<tr>
<th>Source</th>
<th>SSQ</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>(\varepsilon_{p,q})</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (S)</td>
<td>106236.55</td>
<td>11</td>
<td>9657.87</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Difficulty (D)</td>
<td>25585.27</td>
<td>1</td>
<td>25585.27</td>
<td>60.93</td>
<td>---</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x D</td>
<td>4619.4</td>
<td>11</td>
<td>419.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of vehicles (V)</td>
<td>58297.46</td>
<td>2</td>
<td>29148.73</td>
<td>163.27</td>
<td>.84</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x V</td>
<td>3927.77</td>
<td>22</td>
<td>178.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Airspeed (A)</td>
<td>18923.31</td>
<td>2</td>
<td>9461.65</td>
<td>84.43</td>
<td>.78</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x A</td>
<td>2465.35</td>
<td>22</td>
<td>112.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V x D</td>
<td>840.55</td>
<td>2</td>
<td>420.28</td>
<td>3.73</td>
<td>.77</td>
<td>= .05</td>
</tr>
<tr>
<td>S x V x D</td>
<td>2472.49</td>
<td>22</td>
<td>112.38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x D</td>
<td>293.64</td>
<td>2</td>
<td>146.82</td>
<td>3.24</td>
<td>.97</td>
<td>= .06</td>
</tr>
<tr>
<td>S x A x D</td>
<td>996.71</td>
<td>22</td>
<td>45.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V x A</td>
<td>1348.51</td>
<td>4</td>
<td>337.13</td>
<td>3.95</td>
<td>.67</td>
<td>= .02</td>
</tr>
<tr>
<td>S x V x A</td>
<td>3755.89</td>
<td>4</td>
<td>85.36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x V x D</td>
<td>947.81</td>
<td>4</td>
<td>236.95</td>
<td>1.98</td>
<td>.53</td>
<td>= .16</td>
</tr>
<tr>
<td>S x A x V x D</td>
<td>5274.33</td>
<td>44</td>
<td>119.87</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


89
# Appendix P

## Kills Accomplishment Score ANOVA table for high difficulty

<table>
<thead>
<tr>
<th>Source</th>
<th>SSQ</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (S)</td>
<td>438.50</td>
<td>11</td>
<td>54.813</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-First order (1)</td>
<td>179.348</td>
<td>1</td>
<td>179.348</td>
<td>18.71</td>
<td>=.001</td>
<td>0.4090</td>
</tr>
<tr>
<td>S x 1</td>
<td>105.45</td>
<td>11</td>
<td>9.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Second (2)</td>
<td>163.632</td>
<td>1</td>
<td>163.632</td>
<td>41.32</td>
<td>&lt;.0001</td>
<td>0.3732</td>
</tr>
<tr>
<td>S x 2</td>
<td>43.56</td>
<td>11</td>
<td>3.96</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Third (3)</td>
<td>55.712</td>
<td>1</td>
<td>55.712</td>
<td>19.95</td>
<td>=.001</td>
<td>0.1271</td>
</tr>
<tr>
<td>S x 3</td>
<td>30.71</td>
<td>11</td>
<td>2.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Fourth (4)</td>
<td>31.954</td>
<td>1</td>
<td>31.954</td>
<td>3.96</td>
<td>=.07</td>
<td>0.0729</td>
</tr>
<tr>
<td>S x 4</td>
<td>88.72</td>
<td>11</td>
<td>8.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Fifth (5)</td>
<td>1.381</td>
<td>1</td>
<td>1.381</td>
<td>.16</td>
<td>=.70</td>
<td>0.0031</td>
</tr>
<tr>
<td>S x 5</td>
<td>97.93</td>
<td>11</td>
<td>8.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Sixth (6)</td>
<td>.495</td>
<td>1</td>
<td>.495</td>
<td>.09</td>
<td>=.77</td>
<td>0.0011</td>
</tr>
<tr>
<td>S x 6</td>
<td>59.05</td>
<td>11</td>
<td>5.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Seventh (7)</td>
<td>5.164</td>
<td>1</td>
<td>5.164</td>
<td>.93</td>
<td>=.36</td>
<td>0.0118</td>
</tr>
<tr>
<td>S x 7</td>
<td>61.36</td>
<td>11</td>
<td>5.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Eighth (8)</td>
<td>.813</td>
<td>1</td>
<td>.813</td>
<td>.09</td>
<td>=.77</td>
<td>0.0019</td>
</tr>
<tr>
<td>S x 8</td>
<td>100.49</td>
<td>11</td>
<td>9.14</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. First order represents Linear, second order represents Quadratic, etc. R² for polynomial equation is the sum of all significant r²'s.
Appendix Q

Piecewise Regression Best-Fitting Estimations for High Difficulty Kills & DMPIs: Splines & knots

**High Difficulty Kills**

<table>
<thead>
<tr>
<th>Number of points before Knot</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total SSQ</td>
<td>36.5417</td>
<td>8.1033</td>
<td>1.9899</td>
<td><strong>1.9431</strong></td>
<td>6.8065</td>
<td>15.1737</td>
<td>20.5365</td>
<td>64.9436</td>
<td>86.9348</td>
<td>112.6258</td>
</tr>
<tr>
<td>Spline 1 (Slope)</td>
<td>-----</td>
<td>0.7291</td>
<td>0.7532</td>
<td><strong>0.7284</strong></td>
<td>0.6574</td>
<td>0.5948</td>
<td>0.5603</td>
<td>0.4670</td>
<td>0.4201</td>
<td>0.3270</td>
</tr>
<tr>
<td>Spline 2 (Mean)</td>
<td>10.86111</td>
<td>11.4895</td>
<td>11.8214</td>
<td><strong>11.9027</strong></td>
<td>11.983</td>
<td>12.1458</td>
<td>12.1666</td>
<td>12.4583</td>
<td>12.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Best-Fitting Parameters. See Eq. (1).  
Slope of spline 1 = .7284, Mean of Spline 2 = c = 11.903, X-value of knot = M = 16.339, $R^2 = 0.9479$

**High Difficulty DMPIs**

<table>
<thead>
<tr>
<th>Number of points before Knot</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total SSQ</td>
<td>2465.075</td>
<td>980.707</td>
<td>370.207</td>
<td><strong>297.542</strong></td>
<td>422.37</td>
<td>749.429</td>
<td>1003.453</td>
<td>2701.511</td>
<td>3501.399</td>
<td>4811.493</td>
</tr>
<tr>
<td>Spline 1 (Slope)</td>
<td>-----</td>
<td>0.8958</td>
<td>0.8589</td>
<td><strong>0.8390</strong></td>
<td>0.7746</td>
<td>0.7034</td>
<td>0.6564</td>
<td>0.5590</td>
<td>0.5121</td>
<td>0.4082</td>
</tr>
<tr>
<td>Spline 2 (Mean)</td>
<td>79.3241</td>
<td>83.8646</td>
<td>87.1786</td>
<td><strong>88.5417</strong></td>
<td>89.450</td>
<td>91.5625</td>
<td>94.5000</td>
<td>97.7500</td>
<td>95.9167</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Best-Fitting Parameters. See Eq. (1).  
Slope of spline 1 = .8390, Mean of Spline 2 = c = 88.542, X-value of knot = M = 105.522, $R^2 = .8815$
Appendix R

Piecewise Regression Best-Fitting Estimations for Low Difficulty Kills & DMPIs: Splines & knots

### Low Difficulty Kills

<table>
<thead>
<tr>
<th>Number of points before Knot</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spline 1 (Slope)</td>
<td>------</td>
<td>0.9653</td>
<td>0.9180</td>
<td>0.9432</td>
<td>0.9396</td>
<td>0.9177</td>
<td>0.8842</td>
<td>0.8626</td>
<td>0.8149</td>
<td>0.7296</td>
<td></td>
</tr>
<tr>
<td>Spline 2 (Mean)</td>
<td>25.167</td>
<td>26.865</td>
<td>28.262</td>
<td>29.722</td>
<td>31.183</td>
<td>31.417</td>
<td>32.417</td>
<td>32.417</td>
<td>33.000</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

Best-Fitting Parameters. See Eq. (1).
Slope of spline 1 = .8842, Mean of Spline 2 = c = 32.417, X-value of knot = M = 36.663, R² = 0.9827

### Low Difficulty DMPIs

<table>
<thead>
<tr>
<th>Number of points before Knot</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSQ</td>
<td>24823.00</td>
<td>15423.27</td>
<td>9867.19</td>
<td>5049.21</td>
<td>1051.69</td>
<td>723.32</td>
<td>248.491</td>
<td>301.774</td>
<td>1152.99</td>
<td>4491.323</td>
<td></td>
</tr>
<tr>
<td>Spline 1 (Slope)</td>
<td>------</td>
<td>0.9711</td>
<td>0.9222</td>
<td>0.9502</td>
<td>0.9473</td>
<td>0.9371</td>
<td>0.9262</td>
<td>0.9077</td>
<td>0.8721</td>
<td>0.7921</td>
<td></td>
</tr>
<tr>
<td>Spline 2 (Mean)</td>
<td>161.324</td>
<td>172.750</td>
<td>182.726</td>
<td>193.472</td>
<td>205.017</td>
<td>209.146</td>
<td>215.667</td>
<td>219.917</td>
<td>224.500</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

Best-Fitting Parameters. See Eq. (1).
Slope of spline 1 = .9262, Mean of Spline 2 = c = 215.667, X-value of knot = M = 232.852, R² = .9910
## Appendix S

### DMPIs Accomplishment Score ANOVA table for High Difficulty

<table>
<thead>
<tr>
<th>Source</th>
<th>SSQ</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (S)</td>
<td>29580.90</td>
<td>11</td>
<td>2689.17</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-First order (1)</td>
<td>18726.390</td>
<td>1</td>
<td>18726.390</td>
<td>49.81</td>
<td>&lt; .0001</td>
<td>0.6331</td>
</tr>
<tr>
<td>S x 1</td>
<td>4135.73</td>
<td>11</td>
<td>375.98</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Second (2)</td>
<td>7373.933</td>
<td>1</td>
<td>7373.933</td>
<td>41.51</td>
<td>&lt; .0001</td>
<td>0.2493</td>
</tr>
<tr>
<td>S x 2</td>
<td>1954.14</td>
<td>11</td>
<td>177.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Third (3)</td>
<td>1372.447</td>
<td>1</td>
<td>1372.447</td>
<td>20.25</td>
<td>= .001</td>
<td>0.0464</td>
</tr>
<tr>
<td>S x 3</td>
<td>745.36</td>
<td>11</td>
<td>67.76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Fourth (4)</td>
<td>1552.352</td>
<td>1</td>
<td>1552.352</td>
<td>10.46</td>
<td>= .008</td>
<td>0.0525</td>
</tr>
<tr>
<td>S x 4</td>
<td>1632.11</td>
<td>11</td>
<td>148.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Fifth (5)</td>
<td>322.499</td>
<td>1</td>
<td>322.499</td>
<td>1.43</td>
<td>= .26</td>
<td>0.0109</td>
</tr>
<tr>
<td>S x 5</td>
<td>2475.29</td>
<td>11</td>
<td>225.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Sixth (6)</td>
<td>137.31</td>
<td>1</td>
<td>137.31</td>
<td>1.17</td>
<td>= .30</td>
<td>0.0046</td>
</tr>
<tr>
<td>S x 6</td>
<td>1286.70</td>
<td>11</td>
<td>116.97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Seventh (7)</td>
<td>62.346</td>
<td>1</td>
<td>62.346</td>
<td>.61</td>
<td>= .45</td>
<td>0.0021</td>
</tr>
<tr>
<td>S x 7</td>
<td>1120.23</td>
<td>11</td>
<td>101.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Eighth (8)</td>
<td>33.626</td>
<td>1</td>
<td>33.626</td>
<td>.15</td>
<td>= .70</td>
<td>0.0011</td>
</tr>
<tr>
<td>S x 8</td>
<td>2386.65</td>
<td>11</td>
<td>216.97</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. First order represents Linear, second order represents Quadratic, etc. $R^2$ for polynomial equation is the sum of all significant $r^2$'s.
Appendix T

Kills Accomplishment Score ANOVA table for Low Difficulty

<table>
<thead>
<tr>
<th>Source</th>
<th>SSQ</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (S)</td>
<td>5811.50</td>
<td>11</td>
<td>528.32</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Max # of Kills-First order (1)</td>
<td>4810.460</td>
<td>1</td>
<td>4810.460</td>
<td>64.65</td>
<td>&lt; .0001</td>
<td>0.83</td>
</tr>
<tr>
<td>S x 1</td>
<td>818.55</td>
<td>11</td>
<td>74.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Second (2)</td>
<td>919.820</td>
<td>1</td>
<td>919.820</td>
<td>125.31</td>
<td>&lt; .0001</td>
<td>0.16</td>
</tr>
<tr>
<td>S x 2</td>
<td>80.74</td>
<td>11</td>
<td>7.34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Third (3)</td>
<td>.007</td>
<td>1</td>
<td>.007</td>
<td>.0008</td>
<td>.98</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x 3</td>
<td>105.07</td>
<td>11</td>
<td>9.55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Fourth (4)</td>
<td>23.016</td>
<td>1</td>
<td>23.016</td>
<td>1.25</td>
<td>.29</td>
<td>0.004</td>
</tr>
<tr>
<td>S x 4</td>
<td>202.56</td>
<td>11</td>
<td>18.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Fifth (5)</td>
<td>.918</td>
<td>1</td>
<td>.918</td>
<td>.102</td>
<td>.76</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x 5</td>
<td>99.04</td>
<td>11</td>
<td>9.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Sixth (6)</td>
<td>10.826</td>
<td>1</td>
<td>10.826</td>
<td>.837</td>
<td>.38</td>
<td>0.0019</td>
</tr>
<tr>
<td>S x 6</td>
<td>142.32</td>
<td>11</td>
<td>12.94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Seventh (7)</td>
<td>.007</td>
<td>1</td>
<td>.007</td>
<td>.001</td>
<td>.97</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>S x 7</td>
<td>44.29</td>
<td>11</td>
<td>4.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Eighth (8)</td>
<td>46.444</td>
<td>1</td>
<td>46.444</td>
<td>4.21</td>
<td>.06</td>
<td>0.0080</td>
</tr>
<tr>
<td>S x 8</td>
<td>107</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. First order represents Linear, second order represents Quadratic, etc. $R^2$ for polynomial equation is the sum of all significant $r^2$ s.
Appendix U

DMPIs Accomplishment Score ANOVA table for Low Difficulty

<table>
<thead>
<tr>
<th>Source</th>
<th>SSQ</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (S)</td>
<td>297876.068</td>
<td>11</td>
<td>27079.64</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-First order (1)</td>
<td>263054.292</td>
<td>1</td>
<td>263054.29</td>
<td>136.02</td>
<td>&lt; .0001</td>
<td>0.8831</td>
</tr>
<tr>
<td>S x 1</td>
<td>21272.72</td>
<td>11</td>
<td>1933.87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Second (2)</td>
<td>33050.47</td>
<td>1</td>
<td>33050.47</td>
<td>87.18</td>
<td>&lt; .0001</td>
<td>0.1109</td>
</tr>
<tr>
<td>S x 2</td>
<td>4170.05</td>
<td>11</td>
<td>379.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Third (3)</td>
<td>472.383</td>
<td>1</td>
<td>472.383</td>
<td>1.34</td>
<td>= .27</td>
<td>0.0016</td>
</tr>
<tr>
<td>S x 3</td>
<td>3881.98</td>
<td>11</td>
<td>352.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Fourth (4)</td>
<td>688.047</td>
<td>1</td>
<td>688.047</td>
<td>1.28</td>
<td>= .28</td>
<td>0.0023</td>
</tr>
<tr>
<td>S x 4</td>
<td>5916.32</td>
<td>11</td>
<td>537.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Fifth (5)</td>
<td>2.680</td>
<td>1</td>
<td>2.680</td>
<td>.01</td>
<td>= .91</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>S x 5</td>
<td>2401.29</td>
<td>11</td>
<td>218.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Sixth (6)</td>
<td>25.681</td>
<td>1</td>
<td>25.681</td>
<td>.11</td>
<td>= .74</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>S x 6</td>
<td>2514.29</td>
<td>11</td>
<td>228.57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Seventh (7)</td>
<td>281.550</td>
<td>1</td>
<td>281.550</td>
<td>3.63</td>
<td>= .08</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>S x 7</td>
<td>851.29</td>
<td>11</td>
<td>77.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Eighth (8)</td>
<td>300.965</td>
<td>1</td>
<td>300.965</td>
<td>1.73</td>
<td>= .21</td>
<td>0.00101</td>
</tr>
<tr>
<td>S x 8</td>
<td>1908.93</td>
<td>11</td>
<td>173.54</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. First order represents Linear, second order represents Quadratic, etc. R² for polynomial equation is the sum of all significant r²'s.
## Appendix V

### SWAT Redline Rating - Accomplishment Score polynomial ANOVA table for High Difficulty

<table>
<thead>
<tr>
<th>Source</th>
<th>SSQ</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (S)</td>
<td>59424.063</td>
<td>11</td>
<td>5402.19</td>
<td>---</td>
<td>---</td>
<td>0.8878</td>
</tr>
<tr>
<td>Max # of Kills-First order (1)</td>
<td>52759.142</td>
<td>1</td>
<td>52759.142</td>
<td>329.02</td>
<td>&lt; .0001</td>
<td>0.8878</td>
</tr>
<tr>
<td>S x 1</td>
<td>1763.88</td>
<td>11</td>
<td>160.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Second (2)</td>
<td>3598.440</td>
<td>1</td>
<td>3598.440</td>
<td>12.63</td>
<td>= .004</td>
<td>0.0606</td>
</tr>
<tr>
<td>S x 2</td>
<td>3132.847</td>
<td>11</td>
<td>284.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Third (3)</td>
<td>743.456</td>
<td>1</td>
<td>743.456</td>
<td>2.70</td>
<td>= .13</td>
<td>0.0125</td>
</tr>
<tr>
<td>S x 3</td>
<td>3030.56</td>
<td>11</td>
<td>275.51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Fourth (4)</td>
<td>1042.491</td>
<td>1</td>
<td>1042.491</td>
<td>8.78</td>
<td>= .01</td>
<td>0.0176</td>
</tr>
<tr>
<td>S x 4</td>
<td>1306.29</td>
<td>11</td>
<td>118.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Fifth (5)</td>
<td>773.744</td>
<td>1</td>
<td>773.744</td>
<td>4.06</td>
<td>= .07</td>
<td>0.0129</td>
</tr>
<tr>
<td>S x 5</td>
<td>2097.44</td>
<td>11</td>
<td>190.68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Sixth (6)</td>
<td>176.318</td>
<td>1</td>
<td>176.318</td>
<td>.62</td>
<td>= .45</td>
<td>0.0030</td>
</tr>
<tr>
<td>S x 6</td>
<td>3112.60</td>
<td>11</td>
<td>282.96</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Seventh (7)</td>
<td>324.302</td>
<td>1</td>
<td>324.302</td>
<td>2.47</td>
<td>= .14</td>
<td>0.0054</td>
</tr>
<tr>
<td>S x 7</td>
<td>1444.57</td>
<td>11</td>
<td>131.32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Eighth (8)</td>
<td>6.170</td>
<td>1</td>
<td>6.170</td>
<td>.02</td>
<td>= .89</td>
<td>0.0001</td>
</tr>
<tr>
<td>S x 8</td>
<td>3582.65</td>
<td>11</td>
<td>325.70</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. First order represents Linear, second order represents Quadratic, etc. R² for polynomial equation is the sum of all significant r²'s.
Appendix W

SWAT Redline Rating - Accomplishment Score polynomial ANOVA table for Low Difficulty

<table>
<thead>
<tr>
<th>Source</th>
<th>SSQ</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (S)</td>
<td>47331.22</td>
<td>11</td>
<td>4302.84</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-First order (1)</td>
<td>43334.946</td>
<td>1</td>
<td>43334.946</td>
<td>189.58</td>
<td>&lt; .0001</td>
<td>0.9156</td>
</tr>
<tr>
<td>S x 1</td>
<td>2514.42</td>
<td>11</td>
<td>228.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Second (2)</td>
<td>601.719</td>
<td>1</td>
<td>601.719</td>
<td>1.47</td>
<td>= .25</td>
<td>0.0127</td>
</tr>
<tr>
<td>S x 2</td>
<td>4514.65</td>
<td>11</td>
<td>410.42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Third (3)</td>
<td>1138.952</td>
<td>1</td>
<td>1138.952</td>
<td>10.28</td>
<td>= .008</td>
<td>0.0240</td>
</tr>
<tr>
<td>S x 3</td>
<td>1218.63</td>
<td>11</td>
<td>110.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Fourth (4)</td>
<td>113.456</td>
<td>1</td>
<td>113.456</td>
<td>.53</td>
<td>= .48</td>
<td>0.0024</td>
</tr>
<tr>
<td>S x 4</td>
<td>2377.19</td>
<td>11</td>
<td>216.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Fifth (5)</td>
<td>1534.417</td>
<td>1</td>
<td>1534.417</td>
<td>20.78</td>
<td>= .001</td>
<td>0.0325</td>
</tr>
<tr>
<td>S x 5</td>
<td>812.27</td>
<td>11</td>
<td>73.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Sixth (6)</td>
<td>466.050</td>
<td>1</td>
<td>466.050</td>
<td>1.79</td>
<td>= .21</td>
<td>0.0098</td>
</tr>
<tr>
<td>S x 6</td>
<td>2866.07</td>
<td>11</td>
<td>260.55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Seventh (7)</td>
<td>135.297</td>
<td>1</td>
<td>135.297</td>
<td>1.78</td>
<td>= .21</td>
<td>0.0029</td>
</tr>
<tr>
<td>S x 7</td>
<td>835.280</td>
<td>11</td>
<td>75.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Eighth (8)</td>
<td>6.381</td>
<td>1</td>
<td>6.381</td>
<td>.03</td>
<td>= .86</td>
<td>0.0001</td>
</tr>
<tr>
<td>S x 8</td>
<td>2286.84</td>
<td>11</td>
<td>207.89</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. First order represents Linear, second order represents Quadratic, etc. R² for polynomial equation is the sum of all significant r²s.
### Appendix X

**TLX-M4 Redline Rating - Accomplishment Score polynomial ANOVA table for High Difficulty**

<table>
<thead>
<tr>
<th>Source</th>
<th>SSQ</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (S)</td>
<td>30074.537</td>
<td>11</td>
<td>2734.05</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-First order (1)</td>
<td>25387.004</td>
<td>1</td>
<td>25387.004</td>
<td>137.81</td>
<td>&lt; .0001</td>
<td>0.8441</td>
</tr>
<tr>
<td>S x 1</td>
<td>2026.38</td>
<td>11</td>
<td>184.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Second (2)</td>
<td>3098.704</td>
<td>1</td>
<td>3098.704</td>
<td>26.68</td>
<td>= .0003</td>
<td>0.1030</td>
</tr>
<tr>
<td>S x 2</td>
<td>1277.71</td>
<td>11</td>
<td>116.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Third (3)</td>
<td>370.468</td>
<td>1</td>
<td>370.468</td>
<td>4.74</td>
<td>= .052</td>
<td>0.0123</td>
</tr>
<tr>
<td>S x 3</td>
<td>860.44</td>
<td>11</td>
<td>78.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Fourth (4)</td>
<td>584.938</td>
<td>1</td>
<td>584.938</td>
<td>9.28</td>
<td>= .01</td>
<td>0.0194</td>
</tr>
<tr>
<td>S x 4</td>
<td>693.20</td>
<td>11</td>
<td>63.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Fifth (5)</td>
<td>400.636</td>
<td>1</td>
<td>400.636</td>
<td>3.89</td>
<td>= .07</td>
<td>0.0133</td>
</tr>
<tr>
<td>S x 5</td>
<td>1131.60</td>
<td>11</td>
<td>102.87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Sixth (6)</td>
<td>50.095</td>
<td>1</td>
<td>50.095</td>
<td>.71</td>
<td>= .42</td>
<td>0.0017</td>
</tr>
<tr>
<td>S x 6</td>
<td>779.83</td>
<td>11</td>
<td>70.89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Seventh (7)</td>
<td>114.560</td>
<td>1</td>
<td>114.560</td>
<td>.99</td>
<td>= .34</td>
<td>0.0038</td>
</tr>
<tr>
<td>S x 7</td>
<td>1274.84</td>
<td>11</td>
<td>115.89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Eighth (8)</td>
<td>68.132</td>
<td>1</td>
<td>68.132</td>
<td>.45</td>
<td>= .51</td>
<td>0.0023</td>
</tr>
<tr>
<td>S x 8</td>
<td></td>
<td>11</td>
<td></td>
<td>107</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. First order represents Linear, second order represents Quadratic, etc. $R^2$ for polynomial equation is the sum of all significant $r^2$'s.
Appendix Y

TLX-M4 Redline Rating - Accomplishment Score ANOVA table for Low Difficulty

<table>
<thead>
<tr>
<th>Source</th>
<th>SSQ</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (S)</td>
<td>50576.749</td>
<td>11</td>
<td>4597.89</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-First order (1)</td>
<td>46275.709</td>
<td>1</td>
<td>46275.709</td>
<td>215.87</td>
<td>&lt; .0001</td>
<td>0.9150</td>
</tr>
<tr>
<td>S x 1</td>
<td>2358.06</td>
<td>11</td>
<td>214.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Second (2)</td>
<td>627.342</td>
<td>1</td>
<td>627.342</td>
<td>4.38</td>
<td>= .06</td>
<td>0.0124</td>
</tr>
<tr>
<td>S x 2</td>
<td>1574.72</td>
<td>11</td>
<td>143.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Third (3)</td>
<td>1264.039</td>
<td>1</td>
<td>1264.039</td>
<td>10.51</td>
<td>= .008</td>
<td>0.0250</td>
</tr>
<tr>
<td>S x 3</td>
<td>1323.46</td>
<td>11</td>
<td>120.31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Fourth (4)</td>
<td>21.707</td>
<td>1</td>
<td>21.707</td>
<td>.22</td>
<td>= .65</td>
<td>0.0004</td>
</tr>
<tr>
<td>S x 4</td>
<td>1103.21</td>
<td>11</td>
<td>100.29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Fifth (5)</td>
<td>641.766</td>
<td>1</td>
<td>641.766</td>
<td>14.10</td>
<td>= .003</td>
<td>0.0127</td>
</tr>
<tr>
<td>S x 5</td>
<td>500.71</td>
<td>11</td>
<td>45.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Sixth (6)</td>
<td>98.234</td>
<td>1</td>
<td>98.234</td>
<td>2.22</td>
<td>= .16</td>
<td>0.0019</td>
</tr>
<tr>
<td>S x 6</td>
<td>487.24</td>
<td>11</td>
<td>44.29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Seventh (7)</td>
<td>1529.466</td>
<td>1</td>
<td>1529.466</td>
<td>20.43</td>
<td>= .001</td>
<td>0.0302</td>
</tr>
<tr>
<td>S x 7</td>
<td>823.55</td>
<td>11</td>
<td>74.87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Eighth (8)</td>
<td>118.486</td>
<td>1</td>
<td>118.486</td>
<td>1.27</td>
<td>= .28</td>
<td>0.0023</td>
</tr>
<tr>
<td>S x 8</td>
<td>1029.93</td>
<td>11</td>
<td>93.63</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. First order represents Linear, second order represents Quadratic, etc. $R^2$ for polynomial equation is the sum of all significant $r^2$'s.
Appendix Z

TLX All Factors Accomplishment Score ANOVA table for High Difficulty

<table>
<thead>
<tr>
<th>Source</th>
<th>SSQ</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (S)</td>
<td>29486.212</td>
<td>11</td>
<td>2680.57</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-First order (1)</td>
<td>25217.335</td>
<td>1</td>
<td>25217.335</td>
<td>157.47</td>
<td>&lt; .0001</td>
<td>0.8552</td>
</tr>
<tr>
<td>S x 1</td>
<td>1761.60</td>
<td>11</td>
<td>160.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Second (2)</td>
<td>3011.432</td>
<td>1</td>
<td>3011.432</td>
<td>38.06</td>
<td>&lt; .0001</td>
<td>0.1021</td>
</tr>
<tr>
<td>S x 2</td>
<td>870.28</td>
<td>11</td>
<td>79.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Third (3)</td>
<td>431.748</td>
<td>1</td>
<td>431.748</td>
<td>7.12</td>
<td>= .02</td>
<td>0.0146</td>
</tr>
<tr>
<td>S x 3</td>
<td>667.01</td>
<td>11</td>
<td>60.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Fourth (4)</td>
<td>366.523</td>
<td>1</td>
<td>366.523</td>
<td>12.51</td>
<td>= .005</td>
<td>0.0124</td>
</tr>
<tr>
<td>S x 4</td>
<td>322.33</td>
<td>11</td>
<td>29.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Fifth (5)</td>
<td>258.029</td>
<td>1</td>
<td>258.029</td>
<td>3.78</td>
<td>= .08</td>
<td>0.0087</td>
</tr>
<tr>
<td>S x 5</td>
<td>751.30</td>
<td>11</td>
<td>68.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Sixth (6)</td>
<td>159.682</td>
<td>1</td>
<td>159.682</td>
<td>2.93</td>
<td>= .11</td>
<td>0.0054</td>
</tr>
<tr>
<td>S x 6</td>
<td>599.29</td>
<td>11</td>
<td>54.48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Seventh (7)</td>
<td>28.906</td>
<td>1</td>
<td>28.906</td>
<td>.26</td>
<td>= .62</td>
<td>0.0001</td>
</tr>
<tr>
<td>S x 7</td>
<td>1202.74</td>
<td>11</td>
<td>109.34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Eighth (8)</td>
<td>12.557</td>
<td>1</td>
<td>12.557</td>
<td>.09</td>
<td>= .77</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>S x 8</td>
<td>1488.83</td>
<td>11</td>
<td>135.35</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. First order represents Linear, second order represents Quadratic, etc. $R^2$ for polynomial equation is the sum of all significant $r^2$'s.
Appendix AA

TLX All Factors Accomplishment Score ANOVA table for Low Difficulty

<table>
<thead>
<tr>
<th>Source</th>
<th>SSQ</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (S)</td>
<td>44186.397</td>
<td>11</td>
<td>4016.95</td>
<td>---</td>
<td></td>
<td>0.9206</td>
</tr>
<tr>
<td>Max # of Kills-First order (1)</td>
<td>40678.311</td>
<td>1</td>
<td>40678.311</td>
<td>203.88</td>
<td>&lt; .0001</td>
<td>0.9206</td>
</tr>
<tr>
<td>S x 1</td>
<td>2194.74</td>
<td>11</td>
<td>199.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Second (2)</td>
<td>397.780</td>
<td>1</td>
<td>397.780</td>
<td>3.26</td>
<td>= .1</td>
<td>0.0090</td>
</tr>
<tr>
<td>S x 2</td>
<td>1341.20</td>
<td>11</td>
<td>121.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Third (3)</td>
<td>1293.830</td>
<td>1</td>
<td>1293.830</td>
<td>12.90</td>
<td>= .004</td>
<td>0.0292</td>
</tr>
<tr>
<td>S x 3</td>
<td>1102.846</td>
<td>11</td>
<td>100.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Fourth (4)</td>
<td>18.748</td>
<td>1</td>
<td>18.748</td>
<td>0.24</td>
<td>= .63</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>S x 4</td>
<td>850.92</td>
<td>11</td>
<td>77.36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Fifth (5)</td>
<td>674.663</td>
<td>1</td>
<td>674.663</td>
<td>26.96</td>
<td>= .0003</td>
<td>0.01527</td>
</tr>
<tr>
<td>S x 5</td>
<td>275.23</td>
<td>11</td>
<td>25.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Sixth (6)</td>
<td>168.043</td>
<td>1</td>
<td>168.043</td>
<td>4.09</td>
<td>= .07</td>
<td>0.00380</td>
</tr>
<tr>
<td>S x 6</td>
<td>452.21</td>
<td>11</td>
<td>41.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Seventh (7)</td>
<td>893.320</td>
<td>1</td>
<td>893.320</td>
<td>18.46</td>
<td>= .001</td>
<td>0.02022</td>
</tr>
<tr>
<td>S x 7</td>
<td>532.27</td>
<td>11</td>
<td>48.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max # of Kills-Eighth (8)</td>
<td>61.70</td>
<td>1</td>
<td>61.70</td>
<td>1.13</td>
<td>= .31</td>
<td>0.00139</td>
</tr>
<tr>
<td>S x 8</td>
<td>600.46</td>
<td>11</td>
<td>54.59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. First order represents Linear, second order represents Quadratic, etc. $R^2$ for polynomial equation is the sum of all significant $r^2$s.
Appendix AB

Summary of SWAT Axiom Tests

I. Independence

1. Time Load: Time independent of Effort and Stress = 0 failures out of 108 tests
2. Effort Load: Effort independent of Time and Stress = 0 failures out of 108 tests
3. Stress Load: Stress independent of Time and Effort = 0 failures out of 108 tests

II. Double Cancellation

1. Double Cancellation in Time x Effort = 0 failures out of 0 tests
2. Double Cancellation in Effort x Stress = 1 failures out of 3 tests
3. Double Cancellation in Stress x Time = 0 failures out of 0 tests

III. Joint Independence

1. Time x Effort independent of Stress = 6 failures out of 108 tests
2. Effort x Stress independent of Time = 8 failures out of 108 tests
3. Stress x Time independent of Effort = 6 failures out of 108 tests

IV. Approximate Relative Importance of Each Factor

1. Time Load = 30.9722%
2. Effort Load = 34.2682%
3. Stress Load = 34.7595%
Appendix AC

Summary of Mean Raw NASA TLX Ratings by Subscale

<table>
<thead>
<tr>
<th>Low Difficulty Scenarios</th>
<th>Max Possible Kills</th>
<th>Mental Demand</th>
<th>Physical Demand</th>
<th>Temporal Demand</th>
<th>Performance</th>
<th>Effort</th>
<th>Frustration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>12.50</td>
<td>6.67</td>
<td>10.42</td>
<td>5.83</td>
<td>12.50</td>
<td>12.92</td>
</tr>
<tr>
<td>3</td>
<td>19</td>
<td>13.75</td>
<td>10.00</td>
<td>18.33</td>
<td>7.50</td>
<td>15.83</td>
<td>19.58</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>22.50</td>
<td>13.75</td>
<td>30.83</td>
<td>11.67</td>
<td>27.92</td>
<td>25.00</td>
</tr>
<tr>
<td>7</td>
<td>24</td>
<td>20.42</td>
<td>14.17</td>
<td>22.50</td>
<td>10.83</td>
<td>21.25</td>
<td>17.92</td>
</tr>
<tr>
<td>9</td>
<td>34</td>
<td>38.33</td>
<td>27.50</td>
<td>42.50</td>
<td>15.00</td>
<td>41.67</td>
<td>28.75</td>
</tr>
<tr>
<td>13</td>
<td>35</td>
<td>47.50</td>
<td>37.08</td>
<td>70.83</td>
<td>43.33</td>
<td>69.17</td>
<td>50.83</td>
</tr>
<tr>
<td>11</td>
<td>40</td>
<td>48.33</td>
<td>33.33</td>
<td>66.25</td>
<td>25.42</td>
<td>57.92</td>
<td>37.08</td>
</tr>
<tr>
<td>15</td>
<td>46</td>
<td>57.50</td>
<td>34.17</td>
<td>72.08</td>
<td>45.42</td>
<td>65.83</td>
<td>47.08</td>
</tr>
<tr>
<td>17</td>
<td>60</td>
<td>61.67</td>
<td>37.08</td>
<td>84.17</td>
<td>56.25</td>
<td>77.50</td>
<td>62.92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>High Difficulty Scenarios</th>
<th>Max Possible Kills</th>
<th>Mental Demand</th>
<th>Physical Demand</th>
<th>Temporal Demand</th>
<th>Performance</th>
<th>Effort</th>
<th>Frustration</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>8</td>
<td>37.92</td>
<td>17.92</td>
<td>32.50</td>
<td>18.33</td>
<td>36.25</td>
<td>34.17</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>38.75</td>
<td>20.42</td>
<td>46.67</td>
<td>39.58</td>
<td>45.00</td>
<td>44.58</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>51.67</td>
<td>25.42</td>
<td>55.83</td>
<td>48.33</td>
<td>57.92</td>
<td>53.75</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>52.50</td>
<td>27.08</td>
<td>51.25</td>
<td>39.17</td>
<td>57.08</td>
<td>45.00</td>
</tr>
<tr>
<td>10</td>
<td>23</td>
<td>61.67</td>
<td>38.75</td>
<td>66.67</td>
<td>57.92</td>
<td>63.33</td>
<td>60.42</td>
</tr>
<tr>
<td>14</td>
<td>25</td>
<td>69.17</td>
<td>39.58</td>
<td>75.42</td>
<td>68.75</td>
<td>74.58</td>
<td>66.67</td>
</tr>
<tr>
<td>16</td>
<td>36</td>
<td>72.50</td>
<td>30.42</td>
<td>76.67</td>
<td>55.83</td>
<td>74.58</td>
<td>65.42</td>
</tr>
<tr>
<td>12</td>
<td>40</td>
<td>82.50</td>
<td>39.17</td>
<td>85.42</td>
<td>65.83</td>
<td>78.33</td>
<td>64.58</td>
</tr>
<tr>
<td>18</td>
<td>60</td>
<td>87.50</td>
<td>41.25</td>
<td>93.33</td>
<td>82.50</td>
<td>88.33</td>
<td>74.17</td>
</tr>
</tbody>
</table>

Note. The sequence of scenarios follows that of Table 4. In the Low Difficulty condition, the Performance Redline estimate occurred after scenario 13. In the High Difficulty condition, the Performance Redline estimate occurred after scenario 8. Also, the Performance and Physical Demand subscales were not used in the TLX-M4.
Appendix AD

Summary of Percent Kills Data

<table>
<thead>
<tr>
<th>Scenario #</th>
<th>1</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>9</th>
<th>11</th>
<th>13</th>
<th>15</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Fixed</td>
<td>96.53</td>
<td>89.91</td>
<td>97.50</td>
<td>93.40</td>
<td>88.97</td>
<td>81.04</td>
<td>81.19</td>
<td>69.20</td>
<td>55.00</td>
</tr>
<tr>
<td>% All</td>
<td>99.17</td>
<td>99.17</td>
<td>97.50</td>
<td>98.54</td>
<td>94.17</td>
<td>81.04</td>
<td>86.11</td>
<td>71.39</td>
<td>55.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario #</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
<th>14</th>
<th>16</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Fixed</td>
<td>72.92</td>
<td>76.39</td>
<td>57.50</td>
<td>70.83</td>
<td>49.28</td>
<td>32.29</td>
<td>48.33</td>
<td>32.18</td>
<td>20.00</td>
</tr>
<tr>
<td>% All</td>
<td>84.17</td>
<td>72.08</td>
<td>57.50</td>
<td>70.42</td>
<td>47.29</td>
<td>32.29</td>
<td>47.36</td>
<td>34.03</td>
<td>20.00</td>
</tr>
</tbody>
</table>

Note. The “% Fixed” cells show the percentage of Kills achieved up to the fixed time interval of nine minutes. The “% All” cells show the percentage of Kills achieved for the entire scenario.