

THE EFFECTS OF WORKLOAD TRANSITIONS
IN A MULTITASKING
ENVIRONMENT

Thesis

Submitted to

The College of Arts and Sciences of the
UNIVERSITY OF DAYTON

in Partial Fulfillment of the Requirements for

The Degree of
Master of Arts in Psychology

By

Margaret Anna Bowers

UNIVERSITY OF DAYTON

Dayton, Ohio

August, 2013

THE EFFECTS OF WORKLOAD TRANSITIONS IN A MULTITASKING
ENVIRONMENT

Name: Bowers, Margaret Anna

APPROVED BY:

F. Thomas Eggemeier, Ph.D.
Faculty Advisor

James C. Christensen, Ph.D.
Committee Member

Susan Davis, Ph.D.
Committee Member

Concurrence:

Carolyn Roecker Phelps, Ph.D.
Chair, Department of Psychology

ABSTRACT

THE EFFECTS OF WORKLOAD TRANSITIONS IN A MULTITASKING ENVIRONMENT

Name: Bowers, Margaret Anna
University of Dayton

Advisor: Dr. F. Thomas Eggemeier

Previous research has found that performance is sometimes affected by transitions in workload. For some tasks and occupations, this type of change in demand can severely impact performance, which makes this a topic of interest for further research. The previous research conducted used a variety of tasks and methods, but few of the results obtained have been consistent. This study sought to determine the effect of workload transitions in a multi-tasking environment, which is an under-represented area in research on this topic. The use of subjective questionnaires to assess perceived workload and task-related stress has also been limited in previous research. Therefore, this study used the Air Force Multi-Attribute Task Battery (AF-MATB), which is a multi-tasking environment, the NASA Task Load Index, which is a measure of subjective workload, and the shortened Dundee Stress State Questionnaire, which is a subjective

measure of task-related stress. During testing, all participants completed AF-MATB trials that transitioned from an easy level to a difficult level or from a difficult level to an easy level. Also, they all completed easy and difficult control trials. Analyses of the performance data principally supported the success of the task difficulty manipulation, with significant differences only occurring between the easy and difficult portions of trials. However, the results of several AF-MATB subtasks indicated that the transition in task difficulty from difficult to easy had a negative impact on performance compared to performance in the easy control condition. The significant differences in two of these performance measures, however, may reflect trends in the data of the easy control condition as opposed to transition-related decrements in the difficult-easy condition, thereby making conclusions about the presence of transition effects in these instances somewhat difficult. The analysis of both the NASA Task Load Index and of the shortened Dundee State Questionnaire did not reveal any significant differences related to workload transitions.

ACKNOWLEDGEMENTS

Very special thanks are due to my advisor, Dr. F. Thomas Eggemeier. His guidance and support were invaluable during this process! I would also like to thank my committee members, Dr. James Christensen and Dr. Susan Davis. Their encouragement and advice contributed greatly to my success in this endeavor! I truly appreciate all of the time and effort put forth by all three of my committee members. Thank you!

I am also grateful to Iris Davis, who played a crucial role in the training and data collection for the study and was integral in all of the processes before, after, and in-between! I would also like to thank Dr. Greg Funke, whose knowledge of statistics and other counsel helped *enormously*. Thank you! Thanks are also due to Sam Klosterman, who stepped in with MATLAB magic when things were falling apart, to William Miller, whose MATB/MATLAB mastery saved the day, and to Justin Estepp, whose advice and support kept me on target. Lastly, I would like to thank my husband, Drew, for always being in my corner.

TABLE OF CONTENTS

ABSTRACT	iii
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	x
LIST OF TABLES	xii
LIST OF ABBREVIATIONS	xiii
CHAPTER 1 INTRODUCTION	1
Studies Employing Forced-Pace Reaction Time Tasks	3
Studies Employing Memory-Free Tasks	7
Studies Employing Vigilance Tasks	11
Studies Employing Subjective Measures of Workload	18
Studies Employing Dual-Tasks	30
Summary	34
CHAPTER 2 METHOD	43
Participants	43
Materials	44

AF-Multi-Attribute Task Battery	44
DSSQ-3 State Questionnaire.....	51
Training.....	51
CHAPTER 3 DATA COLLECTION	56
CHAPTER 4 RESULTS	58
Performance as a Function of Conditions and Trials.....	61
Performance as a Function of Conditions and Minutes	64
Systems task: mean percent correct responses.....	66
Systems task: mean percent timeouts	68
Systems task: mean percent false alarms	70
Systems task: mean correct reaction time.....	73
Systems task: conclusions.....	76
Tracking: mean RMS error	76
Resource management task: mean deviation from target	78
Communications task: mean percent timeouts and mean correct reaction time.....	80
Conclusions for performance measures	85
Subjective Measures	86
CHAPTER 5 DISCUSSION.....	96
Main Findings	96
Subjective Measures	97

Performance Measures.....	100
Conclusion	108
REFERENCES	111
APPENDIX A: Material for Recruitment.....	119
APPENDIX B: NASA-Task Load Index.....	121
APPENDIX C: DSSQ-3 State Questionnaire	126
APPENDIX D: Consent forms	130
APPENDIX E: Participant Information Form	137
APPENDIX F: Means and standard errors	139
APPENDIX G: ANOVA Tables.....	147
APPENDIX H: Result of Tukey’s HSD tests for percent correct in the Systems task..	155
APPENDIX I: Result of Tukey’s HSD tests for percent timeouts in the Systems task	158
APPENDIX J: Results of Tukey’s HSD tests for mean percent false alarms in the Systems task.....	161
APPENDIX K: Results of Tukey’s HSD tests for mean correct reaction time in the Systems task.....	164
APPENDIX L: Results of Tukey’s HSD tests for mean RMS error in the Tracking task	167
APPENDIX M: Results of the Tukey’s HSD tests for mean deviation from target in the Resource Management task	170

APPENDIX N: Results of the Tukey’s HSD tests for mean percent timeouts in the Communications task.....	173
APPENDIX O: Results of the Tukey’s HSD tests for mean reaction time in the Communications task.....	175
APPENDIX P: Results of Tukey’s HSD tests for mean TLX subscale scores for each condition	178
CURRICULUM VITAE.....	181

LIST OF FIGURES

Figure 1. The Air Force Multi-Attribute Task Battery (MATB) is comprised of four tasks: System Monitoring, Tracking, Communications, and Resource Management.	46
Figure 2. Mean Tracking root mean square error by trial. The error bars represent standard errors.	63
Figure 3. Mean deviation from target in the Resource Management task by trial. The error bars represent standard errors.	64
Figure 4. Mean percent correct in the Systems task by condition and minute. The error bars represent standard errors.	68
Figure 5. Mean percent timeouts in the Systems task by condition and minute. The error bars represent standard errors.	69
Figure 6. Mean percent false alarms in the Systems task by condition and minute. The error bars represent standard errors.	71
Figure 7. Mean correct reaction time in the Systems task by condition and minute. The error bars represent standard errors.	74
Figure 8. Mean root mean square (RMS) error for the Tracking task by condition and minute. The error bars represent standard errors.	78

Figure 9. Mean deviation from target in the Resource Management task by condition and minute. Error bars represent standard errors.	80
Figure 10. Mean percent timeouts in the Communications task by condition and minute. The error bars represent standard errors.	82
Figure 11. Mean correct reaction time in the Communications task by condition and minute. The error bars represent standard errors.	83
Figure 12. Mean, unweighted composites workload scores for NASA TLX in each condition. Error bars represent standard errors.	87
Figure 13. Mean, unweighted composite scores for each TLX subtask in each difficulty condition. Error bars represent standard errors.	89
Figure 14. Average differences scores (z-scores) of distress and engagement from the shortened DSSQ as a function of condition. Error bars represent standard errors.	92
Figure 15. Average differences scores (z-scores) of worry from the shortened DSSQ as a function of condition. Error bars represent standard errors.	93

LIST OF TABLES

Table 1 Rate at which events occur in each level of task difficulty	49
---	----

LIST OF ABBREVIATIONS

Analyses of variance (ANOVA)

Dundee Stress State Questionnaire (DSSQ)

Air Force Multi-Attribute Task Battery (AF-MATB)

NASA Task Load Index (NASA TLX)

Root mean square (RMS)

CHAPTER 1

INTRODUCTION

A workload transition is a shift in the difficulty of a cognitive task, which is of interest because there is evidence supporting the idea that a performer's effectiveness and accuracy may be affected by these transitions. Morgan and Hancock (2011) use driving as an example and point out that, "...the history of previously experienced events may be as influential on driver response and levels of workload as are current levels of demand. The term most applicable for the ongoing influence of such prior historical influences is *hysteresis*" (p. 76). Any effects of prior task workload on post transition workload or performance represent so-called hysteresis effects. For many professions, a transient change in effectiveness is not cause for much concern; however, in some professions, there is little room for error. For instance, Farrell (1999) noted that, "an investigation of air traffic control (ATC) operational errors showed that a high proportion of near misses occurred after a period of sustained high workload, suggesting that the hysteresis effect

may have been a strong contributor” (p. 227). Therefore, the air traffic controllers seemed more likely to make mistakes during a period of low workload if it had been preceded by a period of high workload. The implications of this example are that workload transitions may have far-reaching effects on our safety. Unfortunately, the research already conducted has not led to any firm conclusions about the exact nature of workload transition effects, which suggests that more research on this topic is needed.

In 1993, the subject of these transitions in task demand became more widely recognized with the publication of the Huey and Wickens report for the National Research Council. In 1988, the Army was considering developing tanks with reduced crew sizes, and as a result, “the U.S. Army Human Engineering Laboratory requested that the Committee on Human Factors conduct a study to provide advice and guidance on the effects of prolonged work underload on the subsequent performance of critical tasks and on approaches that could be employed to offset or compensate for decrements in performance that otherwise might occur” (Huey & Wickens, 1993, p. 15). In the report, the authors explore the ways in which this reduction in crew might exacerbate other problems associated with being an active member of a tank crew, which could lead to further problems in its battle readiness. Already, members of tank crews must contend with high levels of workload, stress, and fatigue, but with fewer members, these pressures

would be magnified. The main concern of the report was that the tank crews would be less effective in transitioning from stand-by (which can last up to 72 hours) to full engagement because of these added pressures. However, the authors point out that some of the same issues affecting tank crews could also be extended to “commercial airline crews, nuclear power plant control room crews, railroad freight train crews, merchant and military ship crews, natural disaster relief teams, emergency medical services crews, and trauma center and emergency room crews” (p. 3).

Studies Employing Forced-Pace Reaction Time Tasks

Although the report for the National Research Council brought more attention to the topic, some research on workload transitions had already been conducted prior to its publication. One such study was performed by Cumming and Croft (1973) and was one of the earliest conducted in this area of research. Cumming and Croft used a forced-pace reaction time task (also known as a shadowing task) where 10 digits were presented aurally to participants beginning with a rate of 0.5 digits a second and increasing linearly to 4.0 digits a second. Subsequently, the rate decreased back to 0.5 digits a second. All of this occurred within a short time-frame (trials were 30, 60, 72 seconds in length). Participants used a keyboard to ‘shadow’ as many of the presented digits as possible; if

the speaker on the recording said 'one', the participant would hit the number one key on the keyboard (for more details on the task, refer to Alluisi, Muller & Fitts, 1957).

Regardless of the trial length, Cumming and Croft found that participants' rate of correct responses was significantly higher during the portions of the task with increasing demand than during the portions with decreasing demand. These results led the experimenters to refer to two different theories as explanations. Their first theory was that the participants' short-term memory capacities were being exceeded, and their performances could not resume normal levels until the buffer, a brief memory store, had cleared. This would explain the lag in the performance levels returning to 'normal' (Chamberlain, 1968). Their second theory was that the participants had failed to recognize the change in the level of demand after the workload transition, and consequently, they did not adjust their strategies for the task. This would have led to the observed performance decrement, which would have been only relieved when the participants recognized the level of demand had changed and adjusted their strategies to match. This theory was called the expectancy effects theory (Gibbs, 1965).

Goldberg and Stewart (1980) conducted a study to test the expectancy effects theory from the Cumming and Croft (1973) experiment. They used the same forced-pace reaction time task that Cumming and Croft had used, except their stimuli were presented

visually, instead of aurally. Additionally, each trial was 35.35 seconds in length. The study included three conditions, two of which were meant to cue the participant as to the level of demand that they would be experiencing. In the first cued condition, the digits would begin by appearing at the far left of the screen. The next digit would appear slightly to the right of the first, and the next would be to the right of that and so on. As the task demand increased, the digits would appear further and further right until, at the height of the demand (4.0 digits per second), the digit would appear at the far right side of the screen. As the demand decreased, the digits would progress in their appearance back towards the left side of the screen. The second cued condition was the reverse of the first condition with the digits beginning at the right side of the screen and moving left in their appearance as the demand increased. The third condition did not provide any cues about the task demand; the digits were fixed in the center of the screen. The results did not lend support to expectancy effects theory, because the participants' rate of correct responses under the increasing demand was significantly higher than that under the decreasing demand in the two cued demand conditions, as well as in the control condition. The experimenters concluded that if expectancy effects were the cause, then the cued conditions should have helped the participants recognize that the demand levels were changing and led to better performance scores during the time when demand was

decreasing. As a result of these findings, Goldberg and Stewart concluded that the underlying mechanism must be short-term memory overload.

Farrell (1999) also carried out a study using a forced-pace reaction time task. In this case, the stimuli were letters that were presented on a computer screen; as they appeared, participants pressed the corresponding letter key on a keyboard. The same linearly increasing-decreasing schedule as used in the previous research was used again. There were four conditions; the first had a visual cue where two rectangles reflected the level of demand by their level of emptiness (lower demand) or fullness (high demand). The second condition used an auditory cue to reflect the level of difficulty. Each time a letter was presented, a beep sounded, so as the rate of presentation increased, the frequency of the beeps also increased, and vice versa. The third condition used both auditory and visual cues, and the fourth condition was a control with no cues. Analysis of the results showed that rate of correct responses during the decreasing portion of the task was significantly worse than that of the increasing portion of the task in all four conditions. There were also no significant differences between the cued and control conditions. Thus, the cues had little impact on performance.

Studies Employing Memory-Free Tasks

Matthews (1986) tested the memory overload hypothesis posited by Cumming and Croft (1973) and Goldberg and Stewart (1980). This was accomplished by introducing a task that had no memory component. In the first of two experiments, Matthews chose a task in which the visual load could be manipulated and which also contained a search element. For this task, participants first had to locate, on a computer screen, a group of numeric strings and determine which string was the target and which strings were distracters. The distracters were strings of nonsense variables (ex. $26 + \#1 > T5$), while the target strings were simple arithmetic expressions (ex. $24 + 14 > 27$). There could be 3, 6, 9, or 12 strings within a group. Finally, participants would examine the target string if one was present and manually indicate if the mathematical statement was true or false.

Matthews (1986) designed the study to include 15 trials in each block, and the blocks all took 2.5 minutes to complete. For all trials within a block, the workload level, determined by the number of strings in the group, remained the same. Therefore, workload was manipulated between blocks, and each condition was made up of 16 blocks. In one condition, the workload level increased to its peak (3 strings in the first block, 6 strings in the second, 9 strings in the third, and 12 strings in the fourth) and then

systematically decreased back down to 3 strings. In the second condition, the level of demand for each block was randomized. For the third condition, there was a schedule of high demand for three blocks (12 strings each) and then there would be one block of low demand (3 strings each). That pattern was repeated four times. Finally, the fourth condition was the reverse of the third condition. There were three blocks of low demand (3 strings) and then one block of high demand (12 strings), with that pattern repeated four times. Additionally, there were four control conditions in which each of the participants completed 16 blocks, each at one of the four levels of demand (3, 6, 9 or 12 strings).

The results indicated that, in the randomized demand condition, when a block with high demand was followed by a block of low demand, the search times in the block of low demand were significantly slower than that of the comparable control. This was only true, however, when a load of 12 or 9 strings (but not 6) was followed by a load of 3 strings. When a block of low demand was followed with a block of high demand, search times were significantly faster in the block of high demand than in comparable control conditions. This was true when a block of 3 strings (but not 6 or 9) was followed by a block of 12 strings. These effects persisted for five to six trials before the search times returned to a level consistent with that seen in the control conditions. Matthews (1986) also found that the accuracy of the individuals in the randomized load condition was

significantly less than that of equivalent controls. Participants in the high-load portion of the randomized load condition had the fastest decision latencies but lowest accuracy, which may be indicative of a speed/accuracy trade-off.

Since the results of the first experiment showed a hysteresis effect, despite the fact that the task was devoid of any memory components, Matthews (1986) concluded that the short-term memory overload theory was not the best explanation for the findings. As a result, Matthews proposed the theory of strategic persistence (Poulton, 1982) to explain the results. Strategic persistence occurs when participants fail to switch to the optimum strategy when task demands change (even though they recognize that the demand has changed). Matthews expected that participants who were transitioned from a high level of workload to a lower level would maintain their high level of responding, despite the change; thus, transitioned participants would perform better after the transition than participants in low-workload control conditions. Likewise, Matthews believed that switching participants from a low level to a high level of workload would cause the low level of responding to persist, which would lead participants in high-workload controls to perform better in comparison to participants in the transition condition.

To test this new theory, Matthews (1986) conducted a second experiment. This experiment was much the same as the first, except that the target location was fixed in the

center of the screen, and the participants had to evaluate 12 numeric strings in each of the 5 conditions. Now, instead of manipulating the number of strings each participant had to examine, the rate of presentation was changed. Targets were presented every 3, 6, or 9 seconds. This study employed conditions similar to those used before. In one condition, the rate of presentation linearly increased and then linearly decreased. There was also a condition in which the rate of presentation was randomized. The last 2 conditions either presented several blocks of high workload (high rate of target presentation) followed by several blocks of low workload (low rate of target presentation) or several blocks of low workload followed by several blocks of high workload. Finally, the participants in the control conditions completed 16 blocks of trials at each of the three rates of target presentation.

Matthews (1986) found that the participants in all four of the variable- load conditions performed significantly faster on the blocks with a 6 second and a 9 second presentation rate (but not a 3 second presentation rate) than their equivalent controls. These results were somewhat unexpected and did little to support the theory of strategic persistence.

Studies Employing Vigilance Tasks

Concurrent with this line of research, a separate but related line of research investigated the effects of workload transitions using vigilance tasks. The first of these studies was conducted by Krulewitz, Warm and Wohl (1975). Krulewitz et al. hypothesized that either the habituation model of vigilance (Mackworth, 1968, Mackworth, 1970a, 1970b) or the expectancy theory (Colquhoun, 1961) would be useful in explaining the workload transition phenomenon. The habituation model predicts that a high frequency of background signals (non-critical signals) in a vigilance task would serve to degrade the participants' ability to detect critical signals by reducing the participants' sensitivity to the critical signals. Therefore, a transition from high to low workload would produce worse results than would a transition from low to high workload. The expectancy theory (in regard to vigilance tasks) states that, during low workload, the probability that a signal will be a critical signal is higher than during high workload conditions. Consequently, participants who are shifted from a low level of workload to a high level will expect more of the signals to be critical signals and perform better as a result.

To test these theories, Krulewitz et al. (1975) used a vigilance task in which participants monitored a screen where a red bar would appear. A critical event occurred

when the bar appeared 30 mm instead of 24 mm from center. This study used 20 minutes of induction and 40 minutes of post-transition performance, broken into two, 20 minute periods. The difficulty of the task was manipulated by increasing the number of background signals presented within the timeframe. There were four conditions. The first condition was one that transitioned from a high number of background signals to a low number of background signals (High-low). The second condition transitioned from a low number to a high number of background signals (Low-high). The last two conditions had continuously high levels (High-high) or continuously low levels (Low-low) of background signals.

During the first 20 minutes post-transition, Krulewitz et al. (1975) did not find significant performance differences between the High-low and Low-low conditions or between the Low-high and High-high workload transition conditions. However, participants in the Low-high condition performed significantly worse than those in the High-high condition in the second period of the post-transition session. These results led the authors to reject both the habituation model of vigilance, since no significant results appeared until 20 minutes after the workload shift, and the expectancy theory, because the significant results were not in the expected direction. In the end, the authors suggest that contrast effects may help explain their results. With contrast effects, the first

condition the participant experienced acts an 'anchor' with which any subsequent changes in workload can be compared. For example, initial experience in a low workload condition will affect adjustment to a subsequent higher workload condition. The theory of contrast effects predicts that those who are transitioned from a low workload level to a high workload level will not perform as well as those who are transitioned from a high workload level to a low workload level. In this context, contrast effects would explain the result found in the second post-transition period (those in the Low-high condition performed significantly worse than those in the High-high condition); however, it is interesting that it took 20 minutes for the contrast effects to begin to operate on the results.

Gluckman, Warm, Dember and Rosa (1993) conducted an experiment to examine the theory of contrast effects, as posited by Krulewitz et al. (1975). Gluckman et al. used a vigilance task requiring an absolute judgment of distance with either a dot display or a line display. In the dot display, there were two vertically oriented parallel lines with a dot between them, 10mm from each line. A critical signal occurred when each line was 12 mm from the dot instead of 10mm. The line display also used a pair of parallel lines that were both 36mm long. A critical signal occurred when the lines both were 3mm longer or shorter than the normal 36mm. The single-task portions of the conditions always used

the dot display. The four conditions included in the study were Single task-Dual task, Dual task-Single task, Single task continuous, and Dual task continuous. For the dual-task portions, each trial switched between the dot display and the line display. The testing session lasted for 40 minutes. The first 20 minutes were induction and the second 20 minutes were the post-transition period. Both 20 minute periods were divided into 10 minute segments for subsequent analyses. Analysis of the results revealed no significant differences in the post-transition performance when the Single task-Dual task was compared to the Dual task continuous condition. However, there were significant performance decrements in the Dual task-Single task condition as compared to the Single task continuous condition, but only in the first 10 minutes of the 20 minute post-transition period. When those results were averaged with the results of the rest of the post-transition period, the significant difference was lost (the Dual-Single scores rose to match those of the Single continuous group). The authors concluded that the theory of contrast effects was an unlikely candidate to explain the results they found.

Helton et al. (2004) conducted a study using an abbreviated vigilance task in which participants were to visually detect the target "O" while background signals consisting of "D"s and backward "D"s were also being presented. The researchers manipulated the level of demand in this task by using high and low signal salience. The

conditions used were: high salience switched to low (High-Low), low salience switched to high (Low-High), a consistent low salience control and a consistent high salience control. The analysis of the results revealed no significant differences when the performance of the participants in the High-Low was compared to the performance of those in the low salience control group. Similarly, no significant differences were found in the performance of Low-High as compared to the high salience control group. However, results of a subjective questionnaire, the Dundee Stress State Questionnaire (DSSQ; Matthews et al., 2002; Matthews et al., 1999), showed that the shifted participants experienced more stress than non-shifted controls. The DSSQ also revealed that the shift from high signal salience to low signal salience reduced the participants' task engagement, and the shift from low signal salience to high increased task engagement.

Helton, Shaw, Warm, Matthews, and Hancock (2008) built upon the work by Helton et al., (2004) with the addition of two conditions. In the new conditions, participants were warned of the transition in signal salience. This change allowed experimenters to examine potential differences in reported stress between participants who were and were not warned of the transition. At the beginning of the vigil, the participants in the warned conditions were told that the signal salience would change at

some point during the vigil. The participants in the unwarned conditions were not told that the transition in signal salience would occur. The Dundee Stress State Questionnaire (Matthews et al., 2002; Matthews et al., 1999) was administered pre-vigil and post-vigil. Analysis of the signal detection rate showed no significant differences in performance between participants in either the warned or unwarned conditions as compared to controls. The results of the DSSQ revealed that there were no significant differences in worry across conditions. Additionally, DSSQ ratings of engagement in the unwarned Low-High signal salience condition were significantly lower than ratings in the high salience control condition. DSSQ ratings of engagement in the unwarned High-Low signal salience condition were significantly lower than ratings in the low salience control condition. However, there were no significant differences in reported engagement between the control conditions and the warned transition conditions. Helton et al. theorized that engagement may have been related to participants' appraisal of the challenge associated with the task, and receiving a warning about the transition may have lessened that perceived challenge, resulting in lower engagement scores. The participants in all of the transition conditions reported significantly higher levels of distress, post-transition, than those in the control conditions. Therefore, any shift in signal salience was

more distressing than the continuous salience levels in the control conditions, and receiving the warning about the transition did not seem to ameliorate the effect.

Cox-Fuenzalida, Swikert, and Hittner (2004) also used a vigilance task to test the hypothesis that there is a relationship between neuroticism and the effects of workload transitions. They posited that the transition itself is detrimental to performance because it is a change, and changes can be particularly stressful for individuals scoring high in neuroticism. As a pretest for neuroticism, Cox-Fuenzalida et al. (2004) employed the Eysenck Personality Inventory (Eysenck, 1967). Participants also performed the Bakan vigilance task (Bakan, 1959) in which a sequence of digits is presented via head phones. Ten target sequences of odd-even-odd number strings were presented during each 3-minute trial. The workload for the task was manipulated by increasing the number of distracter digits presented and the speed at which the numbers were presented within the 3-minute trial. The high workload condition contained 225 distracter digits with 1 digit presented every 0.8 seconds, and the low workload condition contained 90 distracter digits with 1 digit presented every 2 seconds. All participants completed 3 low difficulty baseline trials and three high difficulty baseline trials. Each trial was 3 minutes in duration. The participants in the High-to-Low transition condition subsequently completed three trials of high difficulty and one trial of low difficulty. Likewise,

participants in the Low-to-High condition completed three difficult trials followed by one low difficulty trial.

Cox-Fuenzalida et al's (2004) analyses focused on the data collected after the transition in task difficulty, which was the last trial in each sequence. Compared to the participants' 5-minute baselines, the reaction times were significant slower following the transition for both the Low-High and High-Low groups for the individuals who scored higher in neuroticism. This pattern of decrements in performance was also true for the correct response data, but only in the Low-High workload condition. There was no significant relationship between neuroticism and correct responses in the High-Low workload condition.

Studies Employing Subjective Measures of Workload

In another study, Cox-Fuenzalida, Beeler and Sohl (2006) investigated the conclusion drawn in a previous study (Cox-Fuenzalida and Angie, 2005) that transitions negatively affect performance, regardless of the direction of the transition. Cox-Fuenzalida et al. (2006) sought to further examine those results by investigating the possibility that one type of transition (high-to-low or low-to-high) might be more detrimental than the other. To accomplish this, a visual Sternberg memory task

(Sternberg, 1969) was employed. During the Sternberg memory task, participants are given a set of letters to memorize. In this case, participants had 5 seconds to memorize a set of six letters. Then, probe letters were presented, one at time, and participants indicated on a keyboard whether each probe letter was among the memorized set. Probe letters were presented every 0.8, 1.4, or 3 seconds to create high, medium and low workload levels. Participants were assigned to either a High-Medium condition or a Low-Medium condition. Trials were 7 minutes long, and the transition in difficulty occurred after 2 minutes. Also, participants completed a 5-minute baseline trial at each of the workload levels they experienced in the transition trial.

Additionally, Cox-Fuenzalida's study (2006) administered the NASA Task Load Index (TLX; Hart & Staveland, 1988) at the end of the transition trial to assess the participants' perceived level of workload. Not surprisingly, the scores from the TLX showed that participants in the High-Medium workload condition had significantly higher overall workload scores than those in the Low-Medium condition.

Cox-Fuenzalida et al. (2006) also analyzed the data from the Sternberg memory task. The percent of correctly identified probes during the Medium level baseline was compared to that obtained during the first minute after the workload transition. The results of the Sternberg memory task indicated that both sudden increases and sudden

decreases in workload were detrimental to performance as compared to baselines. When the mean performance from the High-Medium trials (first minute after the transition) was compared to the means of the corresponding baseline trials, the difference between the two was larger than when the Low-Medium trials were compared to the corresponding baseline trials. This difference led the authors to conclude that sudden decreases in demand may be more damaging than sudden increases in demand.

Hancock, Williams, Manning, and Miyake (1995) performed a study that focused on how the participants' perception of current workload might be influenced by previous levels of workload. The experiment used a manual, compensatory tracking task, and each condition was made up of three, 5-minute segments. The levels of difficulty for the tracking task in the first condition were Medium-Low-Medium. In the second condition, the levels were Medium-High-Medium, and in the third or control condition, they were Medium-Medium-Medium. Participants completed all three conditions, each on a separate day. The only significant result in the performance data was a decrease in the combined time lead in the last segment of the Medium-High-Medium segment compared to the last segment of the control condition. This indicated that the participants' mean response time became slower in the transition condition.

Hancock et al. (1995) analyzed the results of the Subject Workload Assessment Technique (SWAT; Reid & Nygren, 1988) and found that participants rated the last segment in the Medium-Low-Medium condition as having a significantly higher workload than the last segment in the Medium-Medium-Medium control condition. The same pattern of results was also found with the NASA TLX. With SWAT, the participants rated the last segment in the Medium-High-Medium condition as having a significantly lower workload than the last segment in the Medium-Medium-Medium control condition. The NASA TLX results for the same comparison were not significant. These results indicated that participants' perceptions of workload were somewhat affected by previous levels of demand.

Moroney, Reising, Biers, and Eggemeier (1993) used a task with some resemblance to the compensatory tracking task employed by Hancock et al. (1995). Using Microsoft's "Flight Simulator" (Microsoft, 1990), participants flew a Cessna 182 single engine propeller aircraft through 10 open gates. The difficulty of the task was manipulated by varying the crosswinds, which could be 2, 12 or 22 knots from a heading of 270 degrees. The goal of the study was to examine the effects of previous workload level on performance in a flight simulation task and on ratings in the NASA TLX workload rating scale. Each participant completed three sessions and flew six, 3-minute

trials during each session. The first three trials, the context trials, were all at one of three levels of difficulty (i.e. low, medium or high). The last three trials, the post-context trials, were all at the medium level of difficulty. The within-subjects, counterbalanced Latin Square design ensured that all of the participants experienced all of the levels of difficulty. After each trial, the participants completed the NASA TLX.

Moroney et al. (1993) found significant differences in the performance and in the TLX scores for the context trials at different levels of difficulty. The performance data, which were based on a composite score that took into account number of gates missed, the distance from the crosshairs in the center of the gate, and the time for completion, was significantly better for low difficulty conditions than for the medium or high difficulty conditions. Performance in the medium difficulty condition was significantly better than in the high difficulty condition for the context trials. However, the difficulty of the context trial had no significant effect on the performance attained during the post-context trials. TLX unweighted composite scores showed no significant differences between the low and medium context conditions. However, ratings in both conditions were significantly lower than the ratings in the high condition. In contrast, TLX scores in the post-context medium difficulty conditions showed no significant differences as a function of context conditions.

Moroney et al. (1993) concluded that the level of difficulty in the context trials had no effect on participants' TLX workload ratings during the post-context trials. One of the potential explanations for this result was that the context trials may not have lasted long enough to establish a context situation. Another potential explanation was the administration of the TLX at the end of each trial served as a "terminal point" which allowed the participants to think about each trial as its own separate task. This may have prevented participants from establishing a context between the trials. The authors point out, however, that Hancock, Williams, Miyake, and Manning (1992, unpublished) found context effects while administering the TLX at the end of each trial with even longer breaks between trials (200 seconds in theirs vs. 20-30 seconds in the Moroney et al. study) which makes this explanation less convincing.

Based on the results obtained in the Moroney et al. (1993) study, Fischer, Moroney and Biers (1994) conducted another study, similar to the first, to see if some procedural changes would result in a different outcome. In this study, participants, again, flew a Cessna 182 single-engine, propeller aircraft in a flight simulation created in Microsoft's "Flight Simulator" (Microsoft, 1990). This time, they flew on an S-Curve flight course, through 21 open gates, and a composite performance score was calculated based on the number of gates missed, the distance from the crosshairs in the center of the

gate and the time for completion. The difficulty of the task was manipulated by varying the crosswinds, which could be 2, 12 or 22 knots from a heading of 270 degrees.

Fischer et al. (1994) implemented a new experimental design in which participants completed only three trials during each of their three sessions. The first trial was the baseline trial which was always performed at a medium level of difficulty. The second trial was the manipulation trial, and this trial could be of low, medium or high level difficulty. The third trial was the context trial, and like the baseline, it was always performed at the medium level of difficulty. Since each participant completed three sessions, each separated by 24 hours, all of the participants experienced all of the combinations of trials (Medium-High-Medium, Medium-Low-Medium, and Medium-Medium-Medium). Additionally, each trial was lengthened from 3 minutes to 5 minutes. The new experimental design and lengthened trials were added to promote establishment of context effects for the participants. Finally, after each trial, the NASA TLX was administered.

Fischer et al. (1994) analyzed the performance data from the baseline trials and the baseline un-weighted composite TLX ratings. There were no significant differences found in either measure. The performance data from the manipulation trial was also examined, and participants performed significantly better in the low condition and in the

medium condition than in the high condition; there were no significant differences in performance between the low and medium conditions. Additionally, the composite scores on the NASA TLX were significantly lower for the low and for the medium conditions as compared to the high condition. There were no significant differences in composite scores between the low and medium conditions. Finally, to determine if the levels of difficulty in the manipulation trial affected performance or workload ratings in the subsequent context trial, mixed ANOVAs were performed. There were no significant differences in either the performance data or the NASA TLX scores. It was concluded that the levels of difficulty chosen for flight simulation task might not have been difficult enough to create carry-over effects in the context trial. Another possibility is that the NASA TLX was not sensitive enough to distinguish workload context effects, as previous work showing workload context effects (e.g. Hancock et al., 1992) had used the SWAT.

Moroney, Warm, and Dember (1995) conducted a study examining the effect of abrupt and gradual changes in demand on performance during low-to-high and high-to-low demand transitions. As in the previously described studies, Moroney et al. also examined subjective workload using the NASA TLX. In this study, the participants performed a vigilance task in which they were required to observe two identically

oriented lines. A critical event occurred when the two lines were not oriented in the same direction. During high demand, there were 30 events per minute, and during the low demand, there were 10 events per minute. The 60 minute vigil was divided into a pre-transition, a transition, and a post-transition phase. Each phase was 20 minutes in length. Each trial contained a transition from high to low workload or from low to high workload, and those transitions could occur abruptly or gradually. During the conditions with a gradual shift in demand, one event per minute was added (low-high condition) or removed (high-low condition) over the 20 minutes span of the transition phase. The six conditions used were as follows: 1. abrupt transition, Low- High demand 2. abrupt transition, High-Low demand 3. gradual transition, Low-High demand 4. gradual transition, High-Low demand 5. non-shifted control, high demand 6. non-shifted control, low demand. The results showed that in the Low-High demand group, the participants in both the gradual and the abrupt transition conditions performed significantly worse than their non-shifted controls after the transition period. The detection scores of those in the gradual transition condition (Low-High) remained low during the post-transition phase, but the scores of the participants in the abrupt transition condition rose to levels similar to the non-shifted controls by the end of the post-transition phase. The participants in the High-Low demand groups, in both the abrupt

and gradual transition conditions, all began the vigil with detection scores below those of the non-shifted controls, and their scores remained low for the rest of the vigil.

In the Moroney et al. (1995) study, the NASA TLX was administered to each participant at the conclusion of the vigil. Despite the fact that participants in the Low-High and High-Low conditions all experienced the same levels of demand during the vigil, there were some significant differences in the participants' ratings on the mental demand subscale of the TLX. In the abrupt shift condition, those shifted from High-to-Low reported significantly higher mental demand than those who were transitioned from Low-to-High. Interestingly, the participants in the gradual transition condition reported the opposite effect. Those in the Low-to-High group reported significantly higher mental demand than those in the High-to-Low demand group. Thus, the results of the NASA TLX provide evidence that participants found the High-Low and Low-High conditions to be quite different in terms of the mental demand placed upon them and depending on the other conditions experienced at the time (but only for the mental demand scale). The performance results also indicate that the abrupt vs. gradual manner of demand transition, in some cases, impacted the subsequent performance. However, it appears that more research in this area is needed.

Morgan and Hancock (2011) conducted a study using a driving simulation task to explore the effects of a sudden increase in workload on performance and subjective workload ratings. Workload was assessed by using a simplified version of the SWAT technique (S-SWAT; Luximon & Goonetilleke, 2001). Participants in this study completed four, 5-minute driving trials. During each trial, the heads-up navigation display would malfunction at approximately 3 minutes and 20 seconds into the drive. To correct the malfunction, the drivers would read a 10-digit alphanumeric code aloud to the experimenter. This was the only significant change in workload during the drive.

Morgan & Hancock (2011) examined performance and subjective workload at three points during the drive: once at one minute and twenty seconds into the drive (low workload), again after the navigation error had been corrected (high workload), and finally, at the end of the trial (low workload). Since the S-SWAT assesses three factors (i.e. time, mental effort, and psychological stress) participants verbally reported their ratings (scale of 0-100) to the experimenter at those points during the drive. The performance measures used were the speed differential between the posted speed (45 mph) and the participant's actual speed, the accelerator pedal actuation, and the brake pedal actuation. The brake and accelerator pedal actuation refers to the percentage, between 0 and 100, that the pedal is displaced; a score of 100 percent would indicate that

the pedal was fully depressed. The authors hypothesized that if a hysteresis effect occurred, it would appear as significant differences in the subjective workload scores or in the performance measures between the first measurement period (low workload) and third measurement period (low workload, post-transition), within a trial.

Morgan and Hancock (2011) calculated the overall mental workload score for the S-SWAT by taking an un-weighted average of the ratings on the subscales of time, mental effort, and psychological stress. These S-SWAT scores indicated that workload within trials significantly increased from the first, pre-transition measurement period to the second, high workload period. There was no significant difference between the second and the third measurement periods, although there was a slight decrease in reported workload. For the individual S-SWAT subscales, there was a significant increase in perceived time demand between all of the measurement periods. Similarly, psychological stress also showed significant increases between the first and second period, and the first and third period. However, there were no significant differences between the second and third periods. Last, mental effort increased significantly from the first period to the second and then dropped significantly between the second and third period. However, the third period score was still significantly higher than that of the first period. Therefore, the significant differences between the first period and the third period

in the S-SWAT overall workload scores and in each of the subscales scores all demonstrate hysteresis effects.

Morgan & Hancock (2011) also analyzed the performance data, which showed that the speed differential was significantly greater at the first period than at the second, and there were no significant differences in the scores between the second and third periods. Participants also sped up after the navigation system error, and their speed remained high until the end of the trial. Additionally, the braking actuation score was significantly higher after the navigation system error than before it, and it also remained elevated at the end of the trial. The only significant result for the accelerator actuation was a decrease in the actuation score that occurred in the period after the navigation system error and the end of the trial. Thus, moderate hysteresis effects were demonstrated with the significant differences between first and third periods in the speed and speed differential scores and in the braking actuation scores. However, no hysteresis effects were demonstrated in the accelerator actuation scores.

Studies Employing Dual-Tasks

Using a dual-task paradigm, Cox-Fuenzalida and Angie (2005) used the Sternberg memory task and a simultaneous auditory vigilance task to investigate workload

transition. The two workload levels used in the Sternberg memory task were low and medium. For the low workload level, probe letters were presented every 2 seconds, and in the medium workload level, probe letters were presented every 0.8 seconds.

Participants were either in a Low-Medium testing condition or a Medium-Low testing condition. In both cases, the beginning portion of the trial lasted 2 minutes before transitioning into the second portion of the trial, which was 5 minutes. For the auditory vigilance task, the participant was to listen for a critical tone; the rate of the auditory task remained the same throughout. Each participant also completed two, 5 minute baseline conditions, one at a low level of workload and one at a medium level of workload.

Cox-Fuenzalida and Angie (2005) compared the percent correct on the Sternberg memory task during the first minute after the workload transition with that attained in the baseline sessions. The number of correct responses was significantly lower in both the Low-Medium and Medium-Low transition conditions as compared with baseline performance levels. Therefore, the authors concluded that workload transitions of any type would be a detriment to performance.

Ungar et al. (2005) also conducted a dual-task experiment to examine workload transition effects. Based on the results of the Gluckman et al. (1993) experiment, Ungar et al. hypothesized that the transition effect noted by Gluckman et al. could be explained

by either the resource theory or the strategic-effort-regulation theory. The resource theory model states that individuals have a limited amount of cognitive resources. Some of these resources are consumed during performance of cognitive tasks and more difficult tasks require the expenditure of more resources. When multiple tasks are being performed concurrently, performance on one or both of those tasks is likely to suffer if more resources are needed than are available. Assuming some "carry over" of resource depletion, a participant in a high-low workload condition would have fewer resources for the low portion of that condition and would perform more poorly than someone in a consistent low workload condition. Likewise, an individual who began with a low workload condition and was transitioned to a high workload condition would probably have a higher performance score than a participant who had to perform at high level for the entire condition. The strategic-effort-regulation theory posits that after experiencing a high level of workload, the participant may underestimate the resources needed for a low level of workload, and therefore, not mobilize enough resources for the task. However, according to this model, a participant who is transitioned from a low level of workload to a high level still may not mobilize enough resources for the task, which would ultimately result in lower performance scores relative to a high workload control.

Participants in the Ungar et al. (2005) research either performed an easy or difficulty manual compensatory tracking task and a vigilance task. For the vigilance task, the participant would monitor two dots near the crosshairs of the tracking task and watch for the critical signal, which occurred when the two dots were not equidistant. During single task performance, the participant would only perform the tracking task. During dual task performance, the participant would simultaneously perform the vigilance and the tracking tasks. This study included four conditions: Single task (easy tracking), Single task (difficult tracking), Dual-to single task transition (easy tracking), and Dual-to single task transition (difficult tracking). Ungar et al. used a between subjects design. Each condition was 32 minutes in length, which included 16 minutes of induction and 16 minutes of post-transition performance. The results showed that the participants in the Single task easy tracking condition performed significantly worse on the tracking task than those in the dual-to single task easy tracking condition. These results support the theory of strategic-effort-regulation. Alternatively, those in the Single task difficult tracking performed significantly better on the tracking task than those in the Dual-to single task difficult tracking condition. These results support the resource theory. Therefore, both theories found some support in the data, but neither theory provided a good explanation for all of the results.

Summary

In conclusion, there have been several lines of research on the topic of workload transitions and hysteresis, and there were a number of tasks used and theories proposed within each. The first branch of research began with Cumming and Croft (1973). This study used a forced-pace reaction time task, as did Goldberg and Stewart (1980) and Farrell (1999). All three studies found that while the level of workload was increasing, performance was significantly better than when the level of workload was decreasing. In their discussion, Cumming and Croft hypothesized that the results they found might have been explained by expectancy effects or possibly by short-term memory overload. Goldberg and Stewart tested both hypotheses, and found support for the short-term memory overload theory but not for the expectancy effects hypothesis (since the transitions affected performance even in cued conditions). Farrell explored the same theories as well and concluded that neither short-term memory overload nor expectancy effects alone could explain the effect.

Also within this first branch of research were two studies conducted by Matthews (1986) using a sustained monitoring task in which visual load was manipulated. The results of his first study indicated that participants who transitioned from low demand to high demand performed significantly better than the controls, and those that transitioned

from high demand to low demand performed significantly worse than the controls. Conversely, the results of his second study that year showed that any workload transition, whether it increased or decreased the level of demand, improved performance. Matthews (1986), who was testing a theory of strategic persistence as well as the short-term memory overload theory, found some support for the theory of strategic persistence in the data from Experiment I. Neither study supported the short-term memory overload theory.

The second line of research, using vigilance tasks, began with Krulewitz, et al. (1975) who hypothesized that the habituation model of vigilance or the expectancy theory might be used to explain the effects of workload transitions in sustained attention tasks. Participants who transitioned from a low level of workload to a high level of workload performed significantly worse than a high-high control, but this effect was restricted to a second transition period. This is the opposite effect that Matthews (1986) found in his first study, and Krulewitz et al. concluded that their data did not support either the habituation model of vigilance or the expectancy theory. They speculated that the results obtained might best be explained by contrast effects.

A study examining the effect of abrupt and gradual changes in demand on performance in a vigilance task during low-to-high and high-to-low transitions was

conducted by Moroney et al. (1995). Detection scores in the vigilance task were significantly worse after both types of transition for participants in the Low-High demand group as compared to the non-shifted high workload controls. The detection scores of the participants in the Low-High abrupt transition group rose to those of the controls during the post-shift phase, but the scores of the gradual transition group remained low. The scores for the High-to-Low group under both abrupt and gradual transitions began lower than their non-shifted control counterparts and remained low for the entire vigil. The TLX, administered after each trial, showed that in the abrupt shift condition, those shifted from High-to-Low demand reported significantly higher mental demand than those who were transitioned Low-to-High. However, those in the Low-to-High group reported significantly higher mental demand than those in the High-to-Low demand group when they experienced a gradual shift. These were the only significant differences reflected in the TLX scores. Gluckman et al. (1993) conducted a study using an absolute-judgment vigilance task that switched between two display formats to test the theory of contrast effects posited by Krulewitz et al. (1975). Gluckman et al. found the same pattern of results as that discovered in Matthews's study (Experiment 1, 1986). Their analysis revealed that the transition from high workload to low workload caused a significant

detriment to performance. However, the theory of contrast effects was not supported by these results.

The Gluckman et al., 1993 results were similar to some of the findings reported by Ungar et al. (2005). Ungar et al. used a manual, compensatory tracking task that was paired with a vigilance task under dual-task performance conditions. Performance for participants who were transitioned from a dual task condition involving concurrent difficult tracking and vigilance performance to the single task condition requiring difficult tracking was significantly worse than that of participants who only performed the difficult tracking task in a control condition. However, the performance for the participants who transitioned from a dual task condition requiring concurrent performance of easy tracking and vigilance tasks to a single task involving easy tracking was significantly better than that of the easy tracking controls. The authors had suggested both the resource theory and the strategic-effort-regulation theory as possible explanations for the transition effect, and in fact, both theories found some support in the results obtained.

Moroney et al. (1993) conducted a study to examine the effects of previous workload level on performance in a flight simulation task and on the subsequent NASA TLX ratings. Participants flew a stimulated Cessna 182 single engine propeller aircraft

through 10 open gates, while difficulty was manipulated with varied crosswinds. During each of the three sessions, participants initially flew three trials of the same level of difficulty, medium, low, or high to establish a context. Subsequently, they flew three medium difficulty post-context trials. There were significant differences in the composite flying scores between all of the levels of difficulty under the context trials, with the low difficulty trials having the best scores and the high difficulty trials having the worst scores. However, when the effects of the context trials on the post context trials were examined, there were no significant differences found. Therefore, the level of difficulty experienced in the context trials had no effect on performance during the post-context trials. The differences in un-weighted composite TLX scores for the low versus the medium levels of difficulty during the context trials were not significant, but the TLX scores in both conditions were significantly lower than the scores obtained during the high difficulty context trials. There were no significant differences in the TLX scores for the post-context trials, indicating that there were no transition effects.

Since there were no significant transition effects obtained in the Moroney et al. (1993) study, Fischer et al. (1994) conducted a second study that incorporated some methodological changes. Participants completed trials flying a simulated Cessna 182 during three sessions. Each session contained three, 5 minute trials where the participant

flew through as many of the 21 gates as possible. The first trial or baseline trial and the last or context trial always represented a medium level of difficulty. The second trial, the manipulation trial, could be low, medium, or high difficulty. Analysis of the performance data revealed that the only significant differences obtained were during the second manipulation trial, where the participants performed better on the low and medium difficulty trials than on the difficult trials. The results of the TLX, which was administered after every trial, showed that the composite scores for the low and medium trials were significantly lower than for the difficult trials. The manipulation trials, again, had no significant effect on perceived workload during the context trial. It was suggested that the levels of difficulty within the flight task might not have been different enough from each other for the transition to affect performance. It was also hypothesized that the TLX might not be sensitive enough to reflect any differences in workload that occurred as a result of transitions.

Morgan and Hancock (2011) conducted a study examining the effects of workload transitions during a driving simulation task. Participants completed four, 5 minute driving trials, during which their heads-up navigation system failed at 3 minutes and 20 seconds into the drive. The expectation of the authors was that hysteretic effects would manifest as a significant increase in subjective workload or in performance

between the first period (low workload) and the third period (low workload, post-transition) within a trial. Hysteresis effects were demonstrated by the significant increases in S-SWAT workload scores and in the S-SWAT subscale scores between the first measurement period and the third measurement period. Moderate hysteresis effects were also demonstrated in the differences in the braking actuation scores and in both the speed and speed differential scores between the first and third measurement periods.

Like Ungar et al. (2005), the experiment conducted by Hancock et al. (1995) also used a compensatory, tracking task. Hancock et al. found a significant decrease in the combined time lead when medium demand was preceded by high demand as compared to controls, indicating that the participants' mean response time became slower. However, they also found learning effects from one day of data collection to the next, which would have an effect on their results. Both NASA-TLX and SWAT demonstrated effects in the transition from the low to medium demand condition, and SWAT also showed effects of transition from the high to medium demand condition, thereby demonstrating some workload transition effects.

Finally, Cox-Fuenzalida and Angie (2005) and Cox-Fuenzalida et al. (2006) used the Sternberg memory task in their studies and found that, regardless of whether workload was increased or decreased, transitions in task demand caused a significant

decrements in performance. Cox-Fuenzalida et al. (2004) used the Bakan vigilance task to test the hypothesis that participants who are more neurotic might be less able to cope with demand transitions, which would negatively affect their performance after such transitions. The results of this study demonstrated that individuals scoring higher in neuroticism had longer reaction times than 'normal' participants following both increases and decreases in workload.

As a collection, the previous research conducted on the topic of workload transitions has produced performance results that encompass almost every outcome that might be expected. This is not entirely surprising considering that, across the research that has been conducted, different studies have utilized different types of tasks, different numbers of trials, and trials of different lengths of time in both pre-transition and post-transition conditions. Despite the range of tasks that have been used, however, there have been relatively few studies that used dual task designs (Cox-Fuenzalida & Angie, 2005; Ungar et al., 2005) or simulation tasks (Fischer et al. 1994; Morgan & Hancock 2011; Moroney et al., 1993) and none that have used multiple task environments involving concurrent performance of more than two tasks. Given the lack of studies that have employed multiple task environments to investigate workload transition effects, the current study used the Air Force Multi-Attribute Task Battery (AF-MATB; Miller, 2010;

Comstock & Arnegard, 1992). The AF-MATB is a multiple task environment, and as an aviation simulation, it has the advantage of simulating a real world environment.

Further, there have been a limited number of studies that have incorporated subjective indices of workload into their procedures (Cox-Fuenzalida et al. 2006; Fischer et al., 1994; Hancock et al., 1995; Morgan & Hancock, 2011; Moroney et al. 1995; Moroney et al., 1993). For that reason, the NASA Task Load (Hart & Staveland, 1988) was administered after each trial to examine the effects of workload transitions on subjective workload. Additionally, only two studies (Helton et al., 2004; Helton et al., 2008) have examined the effect of transitions in task demand on subjective stress levels. As a result, this study used the shortened Dundee Stress State Questionnaire (DSSQ-3 State Questionnaire; Matthews, Emo, & Funke, 2005) to assess stress during workload transitions.

Given the history of the prior research on this topic and the inconsistency in results obtained across previous studies, this study was intended to be exploratory in nature. Accordingly, the current study was designed to investigate the effect of transitions in demand on performance, subjective workload, and perceived stress in a multi-task performance battery.

CHAPTER 2

METHOD

Participants

Sixteen participants, 11 men and 5 women, took part in the study. Their ages ranged from 18 to 28 years, with a mean age of 22. All participants were compensated for their participation with \$15 an hour, in the form of Visa[®] gift cards, through Ball Aerospace, Dayton, OH. For recruiting purposes, a mass email (Appendix A) was sent through the Wright-Patterson Public Affairs Office to the Air Force Institute of Technology students, Wright State University students, University of Dayton students and Wright Site Junior Force Council members. The study took place at Wright-Patterson Air Force Base, Dayton, OH, using the facilities of the Adaptive Interface Tech Area (AITA; Building 840, E200L).

Materials

AF-Multi-Attribute Task Battery

This study employed the Air Force Multi-Attribute Task Battery (AF-MATB; Miller, 2010), which was adapted from the original version of the MATB (Comstock & Arnegard, 1992) to be compatible with current computer hardware. Performance data from the AF-MATB, including error rates and reaction times for all of the subtasks, were automatically generated from the simulation software. The AF-MATB was run on a Micron Personal Computer (MPC) Client Pro 565, but the task was mirrored to a standard LCD 19 inch monitor, situated 18 inches in front of the participant. Participants interacted with the AF-MATB tasks through a keyboard, mouse and a Logitech, Extreme 360 Pro joystick. During data collection, an electronic version of the NASA TLX was used for subjective self-assessment of each participant's workload (Appendix B), and the shortened version of Dundee Stress State Questionnaire was used for subjective self-assessment of stress (Appendix C).

The AF-MATB is a PC-based aviation simulation that requires an operator to perform four subtasks simultaneously. These subtasks include: the Systems Monitoring task, the Tracking task, the Resource Management task, and the Communications task

(Figure 1). In the first subtask, the Systems Monitoring task, the participant monitored two warning lights (labeled F5 and F6) and four dials (labeled F1-F4) for errors. The F5 light was normally green, and the F6 light was usually off (or black). If the F5 light turned off, the participant would press the F5 function key on the keyboard to correct the error, and the green light would again turn on. If the F6 light turned red, the participant would press the F6 function on the keyboard to correct the error which would turn the red light off. Additionally, the four dials in the Systems Monitoring task were named F1, F2, F3, and F4, and the indicator in each dial fluctuated up and down. It was acceptable for the indicator to move one tick mark above the middle line and one tick mark below the middle line. If the indicator went outside of that range, the participant would press the function key that corresponded to the dial that was malfunctioning to correct the error.

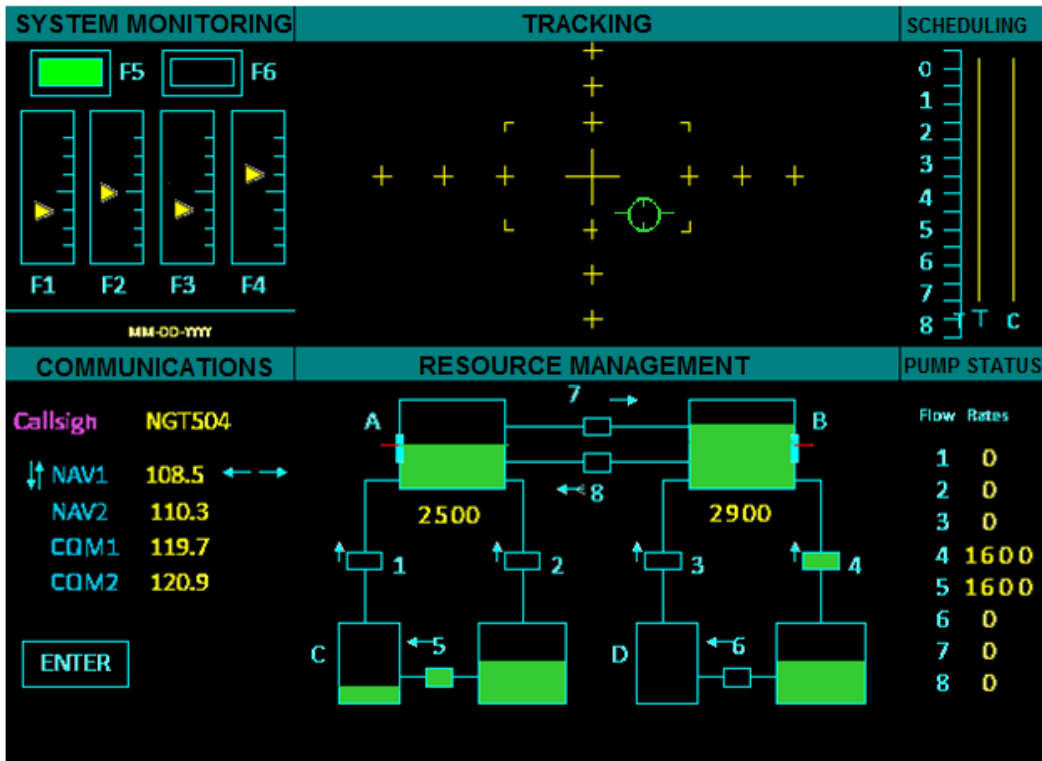


Figure 1. The Air Force Multi-Attribute Task Battery (MATB) is comprised of four tasks: System Monitoring, Tracking, Communications, and Resource Management.

The second subtask was a manual compensatory tracking task. The participant used the Logitech joystick to try to keep the tracking cursor (green circle) as close to the center crosshairs as possible. As this is an unstable tracking task, the tracking cursor moved randomly, and the participant had to compensate for these movements with the joystick.

The Resource Management task was the third subtask. For this task, the participant tried to maintain the fuel levels in the main tanks, Tank A and Tank B, by pumping fuel into them from the subsidiary tanks. The objective was to keep both main tanks at 2500 units of fuel. Since those tanks would lose 2000 units of fuel per minute, Tank C and Tank D were needed to maintain target levels; this was accomplished using Pump 1 and Pump 4. These pumps could move 1800 units of fuel per minute into the main tanks. Additionally, the two unnamed tanks were both 'bottomless' and could pump 1600 units of fuel per minute. The flow rate of each of the pumps was available in the Pump Status window, which was to the right of the Resource Management window. The unlimited tank to the right of Tank C could be used to fill Tank C or Tank A, via Pump 5 or Pump 2 respectively. Likewise, the unlimited tank to the right of Tank D could be used to fill Tank D or Tank B, via Pump 6 or Pump 4 respectively. Also, the participant could use Pump 7 and Pump 8 to move fuel between the two main tanks. A final challenge for this subtask occurred when a pump failed, indicated by a red light, or spontaneously turned off. When a pump failed, there weren't any actions that a participant could take to correct the failure. In the case of a pump failure, the best strategy was usually to utilize other, working, pumps to re-route the fuel from the auxiliary tanks to Tanks A and B. After 10 seconds, the failed pump would correct itself,

which was indicated by the red light turning off. Then, that pump would be available for use again. On the other hand, if a participant noticed that the pump had simply turned itself off, it could be reactivated immediately.

The last subtask was the Communications task. In this task, participants would listen to an audio recording of a male speaker issuing directions to various individuals represented by call signs. The participants would listen for their call sign (e.g. NGT504) and adjust the named channel (e.g., Navigation 1, Navigation 2, Communications 1, Communications 2) to the radio frequencies that the speaker specified.

The overall difficulty of the AF-MATB was manipulated by adjusting the rate of events that occurred during the trial and by adjusting the speed and the number of directional changes of the tracking task. Table 1 shows the event rates that constituted each level of difficulty.

Table 1

Rate at which events occur in each level of task difficulty

Stage	Sequence Length		Communications		Tracking	System Monitoring		Fuel Management	
	Seconds	Minutes	Target	Distractor	Difficulty	Lights	Gauges	Failures	Shut-Offs
Level 1	360	6	4	2	Easy	28	24	2	2
Level 2	360	6	4	4	Easy	34	30	4	2
Level 3	360	6	6	4	Easy	40	36	6	2
Level 4	360	6	6	6	Easy	44	42	6	4
Level 5	360	6	8	6	Moderate	48	48	8	6
Level 6	360	6	8	6	Moderate	54	54	10	6
Level 7	360	6	8	8	Moderate	58	60	12	6
Level 8	360	6	10	8	Moderate	64	66	14	8
Level 9	360	6	10	10	Moderate	70	72	16	8
Level 10	360	6	12	10	Moderate	74	78	18	8
Level 11	360	6	14	10	Moderate	80	82	20	10
Level 12	360	6	16	8	Moderate	82	90	22	10
Level 13	360	6	18	6	High	86	96	24	12
Level 14	360	6	20	6	High	90	102	26	12
Level 15	360	6	20	6	High	96	108	28	14
Level 16	360	6	22	6	High	100	114	30	14
Level 17	360	6	22	4	High	106	120	32	14
Level 18	360	6	24	4	High	110	126	34	16
Level 19	360	6	26	2	High	114	132	36	16
Level 20	360	6	26	2	High	120	138	38	18

NASA Task Load Index.

The NASA TLX (Appendix B) is a subjective measure of workload (Hart & Staveland, 1988) that is comprised of six subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration. After completing a task, participants rate the impact of each subscale on their perceived workload using a scale of 0 to 100, with smaller numbers indicating lower levels of workload and better self-reported performance. After completing the subscale ratings, participants complete the

Sources of Workload, during which, subscales are presented in pairs, and participants choose which subscale had a more significant impact on their perceived workload. Those choices are then used to weight their subscale ratings, and the weighted scores are used to produce a composite score for workload. However, research has shown that unweighted composite scores produce results analogous to the weighted composite scores, and therefore, some experimenters opt not to administer the Sources of Workload (Christ et al., 1993; Hendy, Hamilton, & Landry, 1993; Nygren, 1991).

The TLX is a validated measure that has been used extensively since its creation. Evaluation of the TLX has found high correlations (.94 or above) between all types of administration (paper, computer & aural), and the measure has demonstrated high test-retest reliability, with a correlation of .83 between original scores and the scores collected four weeks later (Hart & Staveland, 1988). However, factor analyses conducted by Bailey, Thompson and Enos (1999) revealed that the five subscales, temporal demand, mental demand, physical demand, effort and frustration, may all contribute to one construct, centering on temporal demands. The sixth subscale, performance, may be a separate construct. Therefore, the idea that the TLX may be measuring multiple constructs should be taken into consideration during evaluation of how the individual subscales contribute to the ratings of overall subjective workload.

DSSQ-3 State Questionnaire

The shortened Dundee Stress State Questionnaire (DSSQ-3 State Questionnaire; Matthews et al., 2005; Appendix C) is a 30 item questionnaire that measures task-related stress, using a scale of 0 to 5. The DSSQ-3 State Questionnaire is based on the 196 item questionnaire called the Dundee Stress State Questionnaire (DSSQ; Matthews et al., 2002; Matthews et al., 1999). The DSSQ-3 State Questionnaire measures three state factors: engagement, distress and worry. In creating the shortened measure, 48 items were initially chosen from the full DSSQ. Of the 48 items evaluated, a subset of eight questions revealed a Cronbach's alpha of .78 for task engagement, and a separate subset of eight questions revealed a Cronbach's alpha of .84 for distress. A final subset of six questions revealed a Cronbach's alpha of .83 for worry (Matthews et al., 2005). Besides those 22 questions, 8 other questions were selected to make up the 30 questions in the DSSQ-3 State Questionnaire. Further validation of the current 30 item measure, such as test-retest reliability, has not yet occurred.

Training

Participants were trained on the task in several sessions, each of which was approximately 2 hours in length. There were at least 5 days of training, and it was not

necessary that those days occurred sequentially. During the first day of training, participants were introduced to the AF-MATB task and to the types of measures that were used during data collection. Then, if they were still interested in participating, they signed consent forms (Appendix D) and completed the participant demographic questionnaire (Appendix E; example questions: age, gender, & handedness). Next, participants began training on the AF-MATB. The participants began with AF-MATB trials that allowed them to practice each of the subtasks individually. There were two trials available for each of the subtasks (i.e. Communications, Systems Monitoring, Tracking, and Resource Management). Then, participants would try as many trials as time would allow, beginning with Level 1. They were also allowed to repeat any level of difficulty they wished.

During the second day of training, the participants completed 20 trials. They began with a trial at Level 1 and worked, sequentially, to Level 20. Because the participants were encouraged to repeat levels with which they had difficulty, many participants did not reach Level 20 until the third or fourth day of training. For the next portion of the training, the participants practiced the higher levels (i.e. 10, 12, 14, 16, 18, & 20) of the task. The participants completed each of the levels three times. At the end of this training session, the experimenter would examine the participants' performance,

and using those data, estimate the level at which each participant consistently reached about 65% performance on the Systems Monitoring task and the Communications task. The participants should also have been fairly consistent across repetitions in the proportion of events that they responded to correctly in those tasks. Finally, their performance on the Tracking and Resource Management tasks should also have been relatively consistent. The level at which the participants achieved these performance criteria was used as their 'estimated titration level.'

The next portion of training was used to ascertain the participants' actual titration level. Titration was a helpful process because participants varied in the degree of success they could achieve with any given level of difficulty in the task, and titration helped ensure that participants were working at levels appropriate to their abilities. To complete the titration session, the experimenter used each participant's 'estimated titration level' as an anchor and used trials at that level of difficulty, trials two levels easier, and trials two levels more difficult with which to test the participants. Therefore, if a participant's 'estimated titration level' was Level 17, s/he was tested on Levels 15, 16, 17, 18, and 19. Participants completed three blocks of the five trials, and the order of the trials within the blocks was randomized. Thus, an example participant's fourth day of training (titration) might look like this:

Block One: Levels 17, 15, 18, 19, 16

Block Two: Levels 15, 19, 16, 17, 18

Block Three: Levels 18, 16, 15, 19, 17

Note that, if a participant's 'estimated titration level' was Level 19 or 20, the levels of difficulty used for titration would be: 16, 17, 18, 19 and 20.

After the participants had completed all 15 titration trials, the experimenter would examine the participants' performance data to determine where they were consistently performing at a level of approximately 65% correct responses in the System Monitoring and Communications tasks. Additionally, the participants' performance on the Tracking task and the Resource Management task should be fairly consistent across repetitions of the chosen titration level.

During the last day of training, participants were introduced to trials that were tailored to their abilities using their titration level. The trials for this final day of training included transitions in task demand. Some trials transitioned from easy to difficult and others transitioned from difficult to easy. The transition in task difficulty occurred after minute 3. Participants also each practiced a trial that was 6 minutes at the easy level (Level 1) and a trial that was 6 minutes at their difficult level (their titrated level). These latter trials were much the same as those experienced during training. The easy trials and

the easy portions of the transition trials were all Level 1 difficulty. The difficult trials and the difficult portions of the transition trials were based on each participant's titration level. Additionally, during the transition trials, participants were asked to push the spacebar if/when they notice a transition in the difficulty of the trial. This button-push served as a check of the participants' subjective experience of the workload. In addition to practicing these AF-MATB trials, the participants were also introduced to the NASA TLX and the shortened DSSQ since both of these measures were presented during data collection. Participants read the instructions associated with each measure and practiced filling out the forms.

CHAPTER 3

DATA COLLECTION

During the data collection session, all of the AF-MATB trials were 6 minutes in length. Participants completed six baseline trials at the easy level (Level 1), six baseline trials at a difficult level (their titrated level of difficulty), and 12 transition trials. Half of the transition trials shifted from difficult to easy, and the other half transitioned from easy to difficult. The easy portions of the transition trials were all at Level 1, and the difficult portions of transition trials were all at the participant's titrated level of difficulty. Each transition in task difficulty occurred after the first 3 minutes of the trial.

All data collection sessions began with the administration of the shortened Dundee Stress State Questionnaire (Matthews et al., 2005) pretest. Next, the participant received one warm-up trial, at an easy level (Level 1), that lasted 6 minutes. Then, the participant completed four blocks of trials. One block of trials had six easy-baseline trials, the second block had six difficult-baseline trials, the third block had six easy-

difficult transition trials, and a fourth block included six difficult-easy transition trials. The order of the blocks of trials was randomized. Participants completed the shortened DSSQ-3 State Questionnaire post-task form following the completion of each block of trials. The DSSQ-3 instructions directed participants to answer according to how they felt while performing each block of trials. Finally, participants completed the TLX at the end of each trial. The participant was instructed to reflect on the second half of the trial while completing the form. The administration of the TLX, along with the several minutes needed for the program to load a new trial, served to create a break (of at least 2 minutes) between the just completed trial and the next trial, which was intended to allow participants to perceive each trial as an independent trial.

CHAPTER 4

RESULTS

Performance on each of the four subtasks was analyzed separately. The System Monitoring task data were based on responses to malfunctions with the gauges and lights, and were analyzed in terms of percent correct responses (hits), percent time outs (misses), and percent false alarms, which were measured as a function of total responses. Correct responses occurred when participants pressed the appropriate button to respond to an event in the System Monitoring task. If they failed to respond to an event, within the 10 second window, it was considered a time out. False alarms occurred when the participant pressed a keyboard button without an associated event occurring. Since there were no background signals to which participants could respond, false alarms were measured in terms of the participants' overall rate of response. Therefore, percent false alarms was calculated by taking the number of false alarms and dividing it by the sum of the number of false alarms and the number of correct responses; that number was then multiplied by

100 to create a percentage $((\text{false alarms}/(\text{false alarms} + \text{correct responses})) * 100)$. Mean reaction time (in seconds) for correct responses was also produced. For the second subtask, Resource Management, performance was measured by the mean deviation from the visible target value (2500 units of fuel) achieved during the trial. Similarly, success in the Tracking task (third subtask) was measured by the root mean square (RMS) error (i.e. the mean deviation of the cursor from center). Finally, performance for the fourth subtask, the Communications task, was assessed in terms of percent correct responses (hits), percent response timeouts (misses), and percent false alarms. In this task, correct responses occurred when the participant followed the audio directions and switched the radio to the correct channel and frequency. Mean reaction time of participants' correct responses measured in seconds was also produced. If participants failed to respond to the instructions issued to their call sign (within the 15 second window), their inactivity was considered a response timeout. Finally, a false alarm occurred when participants responded correctly to directions issued to a call sign other than their own. Interestingly, none of the participants in any of the conditions had any false alarms in the Communications task, so no analysis of this measure took place.

Before conducting any analyses on the data, they were examined for outliers and missing data. Due to random computer errors in the AF-MATB program, performance for a single trial (difficult-easy, Trial 6) was not recorded for one participant. In lieu of

the missing scores, the mean of all of the participants' scores for difficult-easy Trial 6 was used as a replacement. For a second set of analyses, performance for each minute of each trial was calculated. In order to generate those scores for the missing trial, the scores from all of the participants for difficult-easy Trial 6 were averaged, according to minute.

Similarly, one participant was not issued any communication commands during Minute 6 of any of the easy trials. Therefore, replacement scores (percent correct, percent timeout, and the mean correct reaction time) were generated for the communications task for this participant. The mean score, across participants, for Minute 6 for each of the easy trials was used to generate a replacement score for each measure. During Minute 6 of the easy trials, none of the participants let any communications timeout, and they responded correctly to all communication commands. Therefore, using those replacements for scores for this participant did not change the distributions for the percent timeouts and percent correct responses.

Finally, outliers in the data were defined as values that were three standard deviations away from the mean of the group (Tabachnick & Fidell, 2007). In the data divided by trial, each trial in each condition served as a group. Likewise, in the data divided by minute, each minute within each trial served as its own group. If an outlier was found, it was removed from the group, and the mean was calculated again. Then, the

new group mean was used to replace the value of the outlier (Tabachnick & Fidell, 2007). Less than 1% of the data divided by trial were replaced, and 2.5% of the data divided by minute were replaced.

Performance as a Function of Conditions and Trials

As a first step in the analysis, the data were tested for repetition effects as a function of condition for each performance measure. Percent false alarms in the Communications task were omitted from this analysis, because none of the participants answered any of the distracter communications in any condition. For the remaining performance measures, 4 (difficulty) x 6 (trials) repeated measures ANOVAs were conducted. These nine ANOVAs all yielded significant main effects of difficulty condition, $F(3, 45) = 3.49$ to 148.09 , $p = .00$ to $.03$, $\eta_p^2 = .723$ to $.908$. This result supported the expectation that the difficulty manipulation would be successful at producing significant differences in performance. Additionally, seven of the ANOVAs revealed no significant main effects of trial, $F(5, 75) = 0.29$ to 1.85 , $p = .16$ to $.83$, $\eta_p^2 = .019$ to $.110$. This result supports the conclusion that there were no significant learning effects or fatigue effects associated with the repeated trials. Finally, none of the ANOVAS revealed significant interactions between condition and trial for any of the

performance measures, $F(15, 225) = 1.39$ to 1.57 , $p = .22$ to $.66$, $\eta_p^2 = .044$ to $.095$ (See Appendix F, Table 1 for trial means and Appendix G, Table 1 for ANOVA results).

The two performance measures that did have significant main effects of trial were tracking RMS error, $F(5, 75) = 9.30$, $p < .01$, $\eta_p^2 = 0.383$, and mean deviation from target in the Resource Management task, $F(5, 75) = 4.23$, $p < .05$, $\eta_p^2 = 0.220$. Figure 2 illustrates the mean RMS errors for the Tracking task as a function of trial. As can be seen in this figure, performance in the tracking task consistently declined from the beginning to the end of the study. To further investigate these effects, Tukey's Honestly Significant Difference (HSD) tests (Tukey, 1977) were used to compare performance averages from the six trials. The tests comparing tracking RMS error indicated that Trial 1, Trial 2, and Trial 3 all had a significantly lower RMS error than Trial 5 and Trial 6 ($p < .01$). Trial 1 also had significantly lower RMS error than Trial 4 ($p < .05$). It is possible that participants were experiencing some fatigue effects that were manifested in the tracking task. This trials effect could add some ambiguity to the interpretation of the results of this subtask as they pertain to transition effects. However, the noted trials effect should not have a large impact on the analysis of minute-by-minute performance data beyond an increase in the variance of each minute.

The mean deviation from target in the Resource Management task as a function of trial is illustrated in Figure 3. The figure shows that there is some variability in the trial

means. In particular, the mean tank deviation for Trial 2 is greater than any other trial mean. Tukey's HSD tests were used to compare trial means and showed that Trial 2 had significantly higher mean tank deviation than Trials 3, 4 and 5 ($p < .05$). While this trials effect may indicate that performance is not entirely stable across repetitions, it should not have a large impact on the minute-by-minute analyses of performance data beyond an increase in the variance associated with performance within each minute.

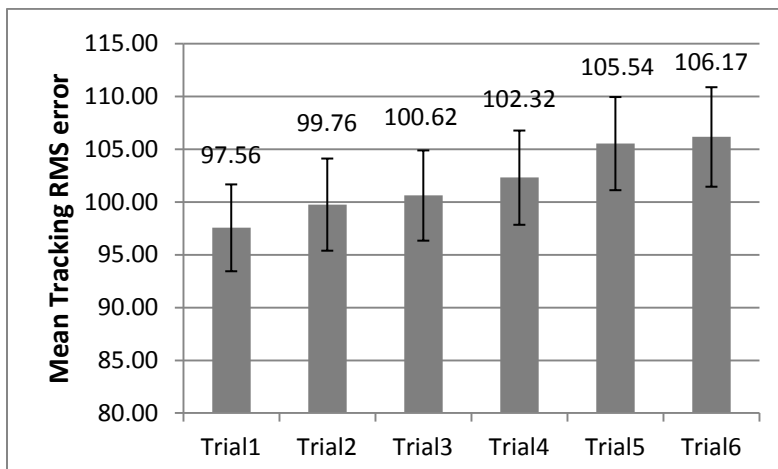


Figure 2. Mean Tracking root mean square error by trial. The error bars represent standard errors.

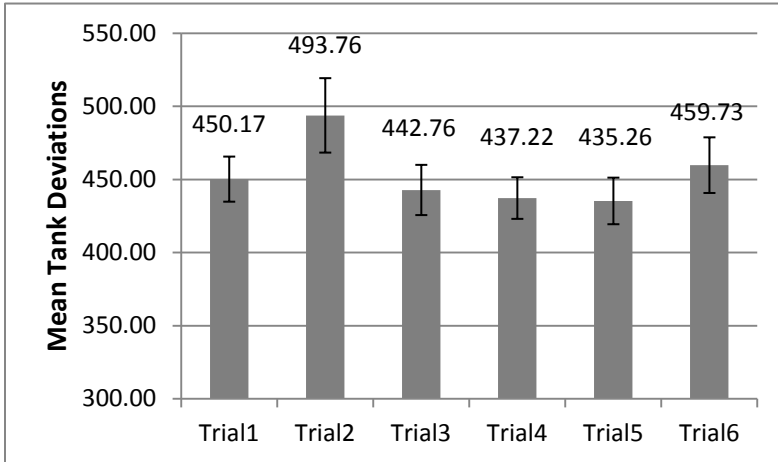


Figure 3. Mean deviation from target in the Resource Management task by trial. The error bars represent standard errors.

Performance as a Function of Conditions and Minutes

The next step in the analysis was to partition performance for each trial into 1-minute increments (for minute-by-minute means, see Appendix F, Table 2). Within each condition, minutes across trials were collapsed to produce minute-by-minute performance means for each participant. Next, 4 (difficulty condition) x 6 (minute) repeated measures ANOVAs were conducted for each performance measure. The results of these nine ANOVAs all showed significant main effects for difficulty condition, $F(3, 45) = 3.35$ to 149.31 , $p = .00$ to $.04$, $\eta_p^2 = .183$ to $.909$. Additionally, several of the ANOVAs showed significant main effects for minute, including percent correct in the Systems task ($F(5, 75) = 5.04$, $p < .01$, $\eta_p^2 = 0.251$) percent timeouts in the Systems task ($F(5, 75) = 5.11$, $p < .01$, $\eta_p^2 = 0.254$) mean correct reaction time in the Communications task ($F(5, 75) =$

10.66, $p < .01$, $\eta_p^2 = 0.135$) and RMS error in the Tracking task ($F(5, 75) = 18.72$, $p < .01$, $\eta_p^2 = 0.555$). Finally, there were significant interactions between condition and minute for eight of the nine ANOVAs, $F(15, 225) = 2.87$ to 166.28 , $p = .00$ to $.03$, $\eta_p^2 = .251$ to $.555$ (For ANOVA results, see Appendix G, Table 2). Percent correct in the Communications task was the only performance measure that did not yield a significant interaction, $F(15, 225) = 2.19$, $p > .05$, $\eta_p^2 = 0.127$. As a result of this non-significant interaction, the performance data for mean percent correct in the Communications task were not analyzed further.

To further investigate the significant interactions for the other eight measures, tests of simple main effects were conducted with one-way ANOVAs to compare performance within each minute as a function of condition. For the measures with significant main effects of difficulty condition in this comparison, Tukey's HSD tests were used to determine which conditions had significantly different means from the others. Because transitions in task difficulty occurred at the beginning of Minute 4 in the transition conditions, performance in Minutes 4, 5, and 6 were of particular interest in regards to the possible effects of workload transitions on performance. Specifically, it was meaningful if performance during Minutes 4, 5, or 6, in the transition scenarios was significantly different from its comparable control. For example, if performance in Minute 4 was significantly different in the difficult-easy condition than in Minute 4 of the

easy condition, it could be theorized that the transition impacted performance. The same would be true if performance was different in Minute 5 of the easy-difficult condition as compared to Minute 5 of the difficult condition, for example. The analysis of each performance measure is detailed below.

Systems task: mean percent correct responses

One-way, repeated measures ANOVAs were conducted for percent correct in the Systems task for each minute, and all were significant for the simple main effect of difficulty condition, $F(3, 45) = 26.83$ to 79.68 , $p = .000$, $\eta_p^2 = .641$ to $.842$ (for ANOVA results, see Appendix G, Table 3). In Figure 4, the mean percent correct in the Systems task is represented as a function of minute and difficulty condition. As can be seen in the figure, the means of the easy and easy-difficult conditions were very similar in Minutes 1-3, and the same was true of the difficult and difficult-easy conditions. At the beginning of Minute 4, the workload manipulation occurred in the transition conditions. For Minutes 4-6, the means of the easy and difficult-easy conditions were very similar, and the means of the difficult and easy-difficult conditions were similar to each other as well. Indeed, the Tukey's HSD tests used to follow up the one-way ANOVAs found that the only significant differences in each minute comparison were between the easy and difficult portions of the trials ($p < .05$) (for results of Tukey's HSD tests, see Appendix

H). For example, during Minutes 1-3, there was a significantly higher mean percent correct in the easy condition than in the difficult-easy and the difficult conditions. Also, performance was significantly better in the easy-difficult condition than in the difficult-easy and the difficult conditions. After the transition occurred (Minutes 4-6), performance in the easy condition was still significantly better than that in the difficult condition, but now, the easy condition was significantly better than the easy-difficult condition. Also, the easy condition was no longer significantly different compared to the difficult-easy condition. Likewise, performance in the difficult-easy condition was better than in the difficult condition and better than in the easy-difficult condition, but was not different from the easy control condition.

In summary, these results show that the only significant differences within each minute comparison were between the easy and difficult portions of the trials. Although the differences found were not unexpected, none supported the position that workload transitions impact performance. Significant differences between the difficult-easy and the easy conditions or between the easy-difficult and the difficult conditions, in Minutes 4, 5 or 6 would have supported the existence of a transition effect. However, no such differences were found for this measure.

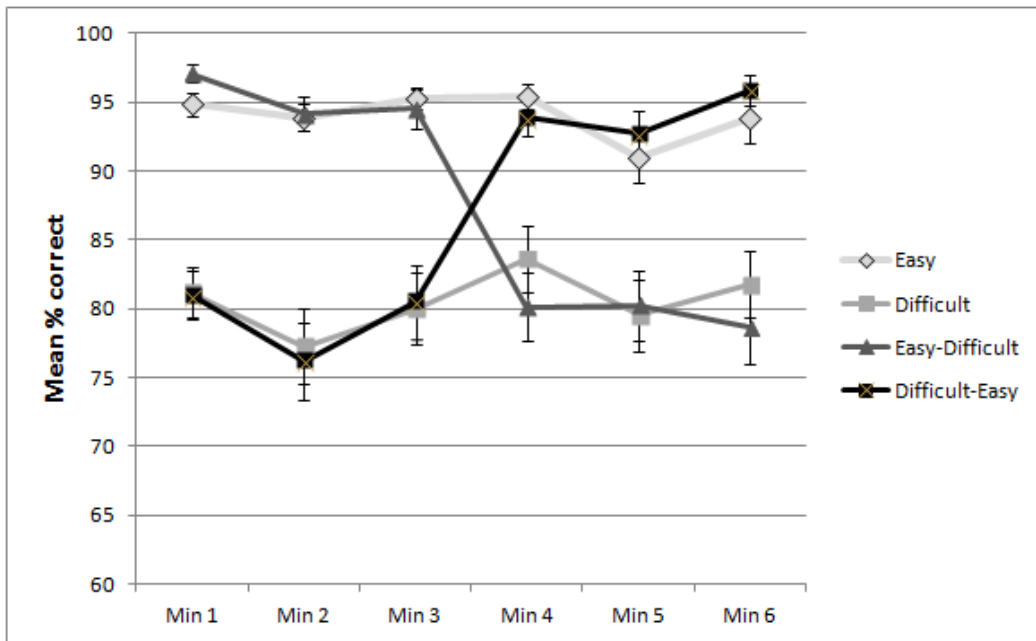


Figure 4. Mean percent correct in the Systems task by condition and minute. The error bars represent standard errors.

Systems task: mean percent timeouts

One-way, repeated measures ANOVAs were also conducted for percent timeouts in the Systems task within each minute, and all of the simple main effects were significant, $F(3, 45) = 26.83$ to 79.68 , $p = .000$, $\eta_p^2 = .641$ to $.842$ (for ANOVA results, see Appendix G, Table 3). Mean percent timeouts in the Systems task, as a function of difficulty condition and minute, are illustrated in Figure 5. As shown in the figure, the percent timeouts in the easy and the easy-difficult conditions were lower than those in the difficult and difficult-easy conditions during Minutes 1-3. After the workload transition

in Minute 4, the percent timeouts in the easy and difficult-easy conditions were lower than those in the easy-difficult condition and the difficult condition. Like the results obtained with the Tukey's HSD tests in the Systems task for percent correct, only the difficult and easy portions of the trials were significantly different ($p < .05$) (for results of Tukey's HSD tests, see Appendix I). Thus, there were no effects of workload transition on performance for this measure.

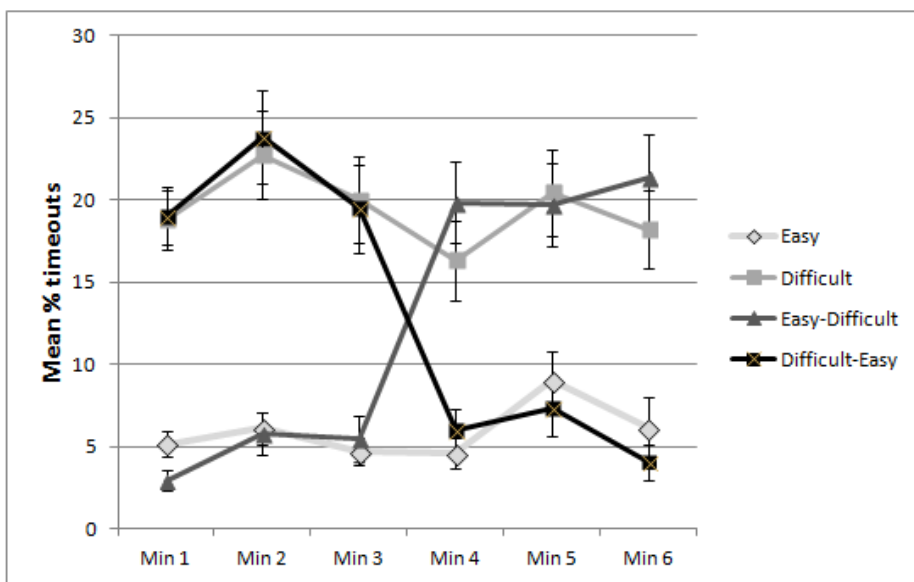


Figure 5. Mean percent timeouts in the Systems task by condition and minute. The error bars represent standard errors.

Systems task: mean percent false alarms

In order to further investigate the significant difficulty condition x minute interaction, one-way, repeated measures ANOVAs were conducted for percent false alarms in the Systems task for each minute. The simple main effects were significant in each instance $F(3, 45) = 5.02$ to 22.05 , $p = .000$ to $.016$, $\eta_p^2 = .251$ to $.595$ (for ANOVA results, see Appendix G, Table 3). Figure 6 illustrates mean percent false alarms as a function of minute and difficulty condition. As shown in the figure, the mean percent false alarms for the easy, difficult, and difficult-easy conditions were all roughly the same during Minute 1. The Tukey's HSD tests for Minute 1 demonstrate that those three conditions were significantly different than the easy-difficult condition. Figure 6 also shows that trends similar to those seen for percent correct and percent timeouts resume with Minutes 2 and 3. Tukey's HSD tests revealed that there was a lower percent false alarms for the easy and easy-difficult conditions as compared to the difficult and difficult-easy conditions for Minute 2 ($p < .01$), Minute 3 ($p < .05$), and Minute 5 ($p < .01$). Analysis of Minute 6 found that the difficult-easy condition had a lower percentage of false alarms than the difficult ($p < .05$) and easy-difficult conditions ($p < .01$). However, there were no significant differences between the easy and difficult or the easy and easy-difficult conditions during Minute 6. Unlike the analysis of any other minute,

the Tukey's HSD tests for Minute 4 revealed significant differences in percent false alarms between the difficult-easy and the easy control conditions (described next).

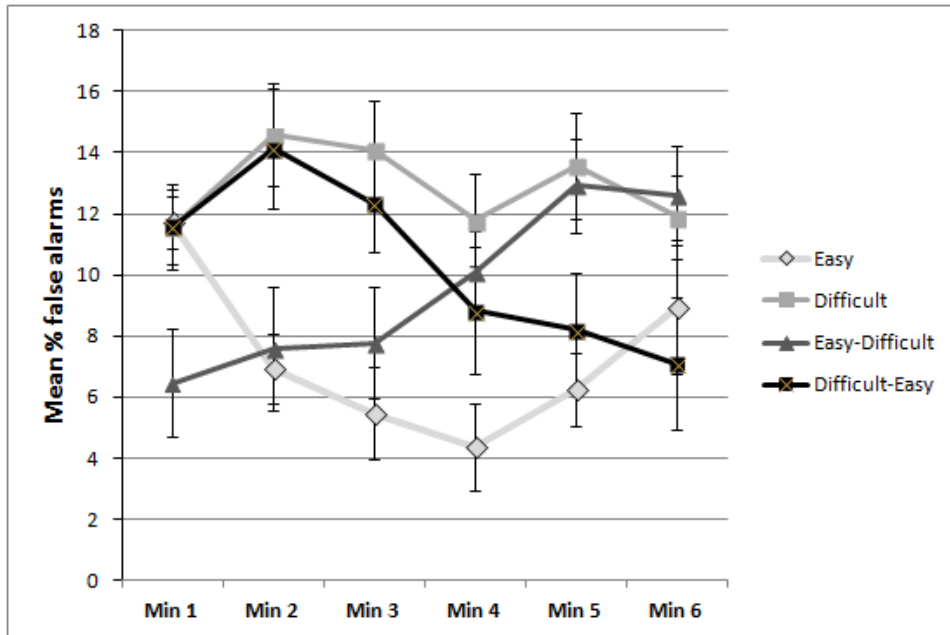


Figure 6. Mean percent false alarms in the Systems task by condition and minute. The error bars represent standard errors.

During Minute 4, Tukey's HSD tests revealed that the easy condition had significantly lower percent false alarms than the difficult-easy condition, which could indicate that the transition from a high level of workload to a low level negatively affected performance compared to the easy control ($p < .05$). However, the performance trends in the easy condition across Minute 1 through 4 show progressive declines in the percentage of false alarms. These declines persisted after percent false alarms in the easy

condition reached a level that approximated those in easy-difficult condition during Minute 2. The persistence of declines across Minutes 3 and 4 was unexpected and suggests that the difference noted between the easy and difficult-easy conditions during Minute 4 may at least partially reflect this decline as opposed to transition-related decrements in the difficult-easy condition.

In order to determine if the mean percent false alarm data from Minute 4 significantly differed from the other minutes in the easy control condition, a one-way ANOVA was conducted. This ANOVA revealed a significant main effect of minute, $F(5, 75) = 6.40, p < .01, \eta_p^2 = 0.299$. The Tukey's HSD tests used to investigate further revealed that mean percent false alarms in the easy control condition during Minute 4 did not differ significantly from Minutes 2, 3 or 5. Therefore, the performance difference noted between the easy and difficult-easy conditions during Minute 4 cannot be fully explained by the declining trends in Minutes 3 and 4 in the easy control condition. Thus, the significant difference between the easy and difficult-easy condition in Minute 4 may at least partially reflect transition-related decrements in the difficult-easy condition (for results of Tukey's HSD tests, see Appendix J).

In summary, the majority of the significant differences between conditions in the percent false alarms measured in the Systems task were between the easy and difficult segments of the trials. However, in Minute 4, there was a significantly higher percent

false alarms in the difficult-easy condition as compared to the easy condition. Although this result might be partially explained by the trends in the data in the easy control condition, performance during Minute 4 did not differ significantly from Minutes 2, 3 or 5 in the easy control condition. Therefore, the difference between the easy-difficult and the easy condition in Minute 4 may reflect transition-related decrements in the difficult-easy condition.

Systems task: mean correct reaction time

To follow up the significant difficulty condition x minute interaction, one-way, repeated measures ANOVAs were conducted for each minute to compare mean correct reaction time (in seconds) across conditions. The ANOVAs revealed significant simple main effects of difficulty condition for each minute, $F(3, 45) = 23.47$ to 67.67 , $p = .000$, $\eta_p^2 = .610$ to $.819$ (for ANOVA results, see Appendix G, Table 3). Figure 7 illustrates the mean reaction times for correct responses in the Systems task as a function of difficulty condition and minute. As the figure shows, the means are generally lower (faster) in the easy portions of the trials as compared to the difficult portions of the trials. The Tukey's HSD tests conducted supported this conclusion with significant differences between all easy and difficult portions of trials (for results of Tukey's HSD tests, see Appendix K). This account of the performance differences was true for all of the minutes except Minute

6. During Minute 6, participants' mean correct reaction time was significantly faster in the easy condition than in the difficult-easy condition ($p < .05$), which might indicate that the transition in workload negatively affected their mean correct reaction time on a delayed basis (i.e., during Minute 6 versus Minutes 4 or 5). However, mean reaction time in the easy condition during Minute 6 declined relative to levels observed during Minutes 4 and 5. This unanticipated decline, coupled with an increase in mean reaction time in the difficult-easy condition, may have contributed to the significant difference between these conditions during Minute 6.

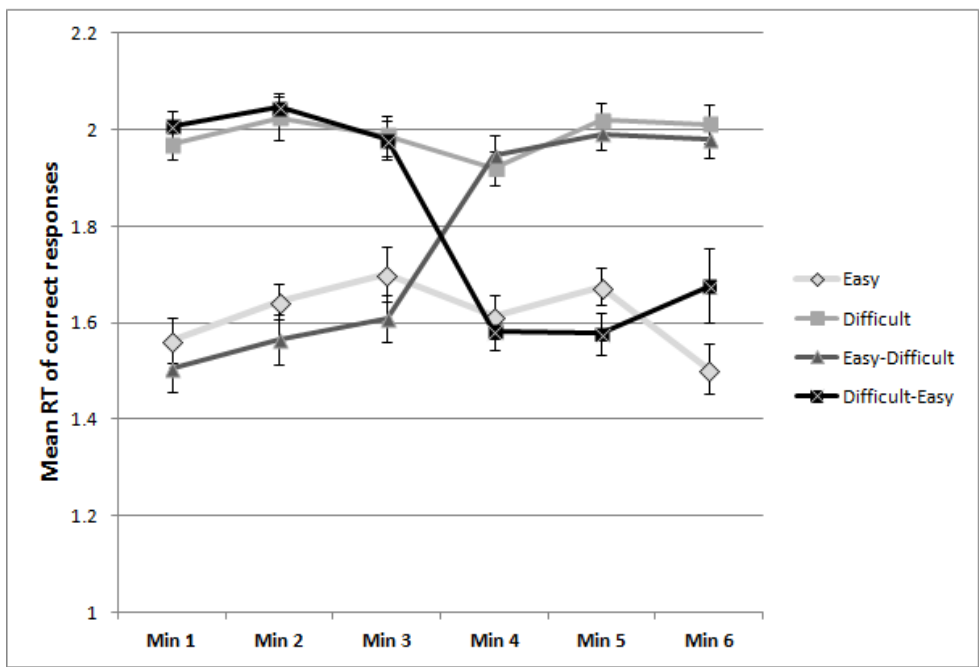


Figure 7. Mean correct reaction time in the Systems task by condition and minute. The error bars represent standard errors.

In order to further investigate performance in the difficult-easy and easy control conditions during Minute 6, one-way ANOVAs were conducted. The one-way ANOVAs revealed a significant main effect of minute for both the easy control condition ($F(5, 75) = 6.40, p < .01, \eta_p^2 = 0.30$) and the difficult-easy condition ($F(5, 75) = 34.62, p < .01, \eta_p^2 = 0.70$). Tukey's HSD tests used to follow up these significant main effects. The Tukey's HSD tests revealed that the only significant differences between minutes in the difficult-easy condition were between the first half (Minutes 1-3) and the second half (Minutes 4-6) of the condition. There were no significant differences in the mean correct reaction time between Minutes 4, 5 and 6. Additionally, further analysis of performance in the easy control condition revealed that the mean correct reaction time during Minute 6 was significantly lower than during Minute 2 ($p < .05$), Minute 3 ($p < .01$), and Minute 5 ($p < .05$). These results suggest the difference between the easy control condition and the difficult-easy condition in Minute 6 may be more likely attributable to the differences in performance in the easy control condition than to transition-related decrements in the difficult-easy condition. However, clear interpretations of the data, in terms of presence or absence of transition-related effects in Minute 6, are difficult.

Systems task: conclusions

In conclusion, most of the Tukey's HSD tests showed that the only significant differences in each minute, observed across conditions, were between the easy and difficult portions of the trials. However, analysis of performance in the Systems task revealed that there was a significantly lower percent false alarms in the easy condition than the difficult-easy condition in Minute 4. Participants also had a significantly slower mean correct reaction time in the difficult-easy condition than in the easy condition in Minute 6. Both of these results potentially support the position that transitions in workload have an impact on performance. However, as noted above, the trends in performance in the easy conditions may have influenced the outcomes of those comparisons.

Tracking: mean RMS error

In order to investigate the significant difficulty condition x minute interaction, one-way repeated measures ANOVAs were conducted within each minute to compare tracking RMS error across conditions. The ANOVAs revealed significant simple main effects of condition for every minute, $F(3, 45) = 136.02$ to 229.63 , $p = .000$, $\eta_p^2 = .901$ to $.939$ (for ANOVA results, see Appendix G, Table 3). The mean RMS error for the Tracking task, as a function of difficulty condition and minute, is presented in Figure 8.

The figure illustrates that, in all cases, the RMS error is lower in the easy portions of the trials as compared to the difficult portions of the trials. This result is generally consistent with the pattern observed previously in the Systems task. Tukey's HSD tests were conducted to follow-up all significant main effects. Significant differences between the easy and difficult portions of the trials were found within each minute ($p < .01$) (for results of Tukey's HSD tests, see Appendix L). Performance in transition Minute 4 was unique, however. The RMS error for the easy condition was significantly lower than in the difficult-easy condition ($p < .05$), which mirrors the results of the percent false alarms in the Systems task. Therefore, this result also lends support to the position that workload transitions affect performance.

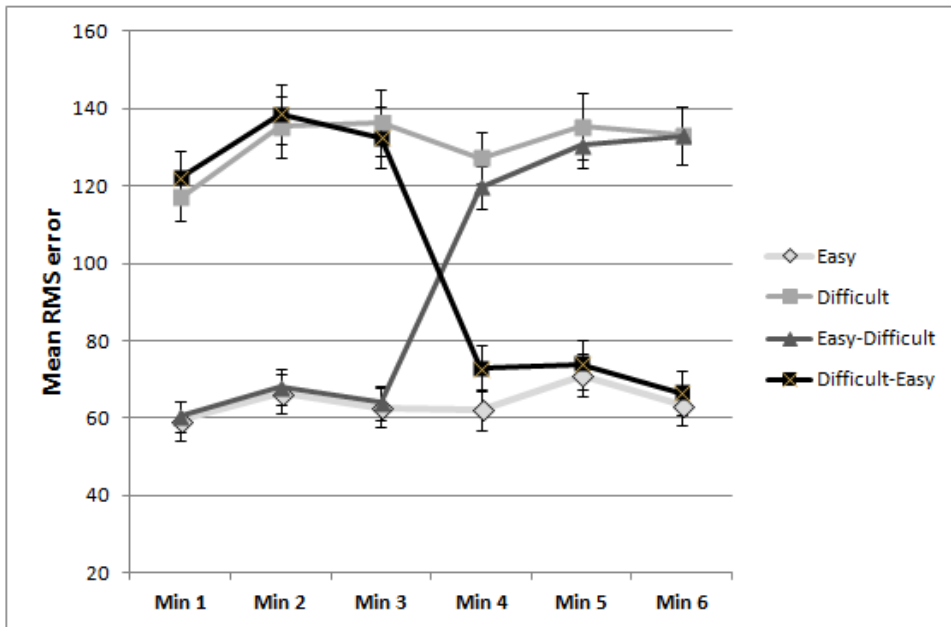


Figure 8. Mean root mean square (RMS) error for the Tracking task by condition and minute. The error bars represent standard errors.

Resource management task: mean deviation from target

To investigate the significant difficulty condition x minute interaction in the Resource Management task, one-way, repeated measures ANOVAs were conducted for the mean deviation from target measure. These ANOVAs revealed significant simple main effects of condition within each minute $F(3, 45) = 3.62$ to 9.61 , $p = .000$ to $.032$, $\eta_p^2 = .194$ to $.390$ (for ANOVA results, see Appendix G, Table 3). Mean deviation from target in the Resource Management task, as a function of condition and minute, is shown

in Figure 9. As can be seen in the figure, the mean deviations are smaller in the easy portions of trials as compared to the difficult portions of the trials, but the error bars show that the standard error for each mean is relatively large. The Tukey's HSD tests, used to follow up the one-way ANOVAs, demonstrated that the differences between easy and difficult portions of trials were not always significant (for results of Tukey's HSD tests, see Appendix M). Although mean performance was ordered as would be expected, a number of differences between easy and difficult trial segments failed to achieve significance. For instance, in Minute 2, the mean deviation for the difficult condition was significantly higher than the mean deviation for the easy and the easy-difficult conditions, but the mean deviation for the difficult-easy condition was not significantly higher than the deviations in the easy condition or the easy-difficult condition. It may be the case that factors, such as shifting task priorities or individual differences in approaching the task, led to the relatively high variability in participant scores that was observed, thereby contributing to this pattern of results. Each minute had at least one significant difference between the easy and difficult portion of the trials, but none of these significant differences supported the position that the workload transitions had an impact on performance.

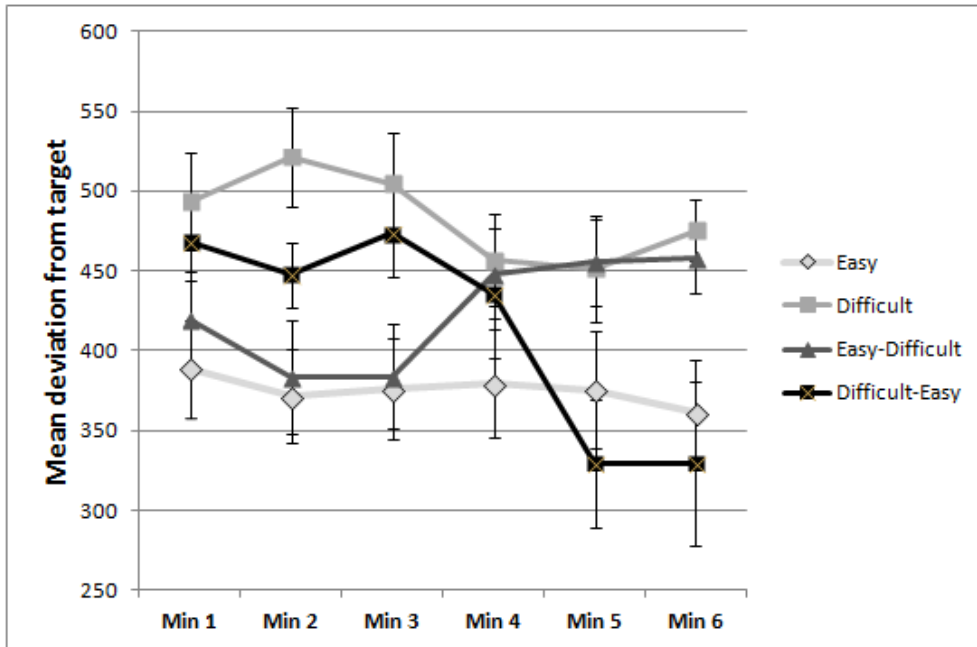


Figure 9. Mean deviation from target in the Resource Management task by condition and minute. Error bars represent standard errors.

Communications task: mean percent timeouts and mean correct reaction time

To investigate the significant difficulty condition x minute interaction in the Communications task, one-way, repeated measures ANOVAs were conducted for the mean percent timeouts and the mean correct reaction time measures (for ANOVA results, see Appendix G, Table 3).

Significant simple main effects of condition were found for percent timeouts only in Minute 3 ($F(3, 45) = 3.56, p < .05, \eta_p^2 = 0.192$) and Minute 6 ($F(3, 45) = 3.98, p < .05, \eta_p^2 = 0.210$). Figure 10 illustrates the mean percent timeouts in the Communications task as a function of difficulty condition and minute. As can be seen in the figure, the trends in the data for Minute 3 and 6 show that, in general, there is a higher percent timeouts in the difficult portions of the trials as compared to the easy portions. The figure also illustrates that in every minute, the mean percent timeouts and the standard errors associated with each mean appear to be somewhat variable. The significant one-way ANOVAs were further investigated using Tukey's HSD tests. The Tukey's HSD tests revealed that during Minute 3, there was a significantly lower percent timeouts in the easy-difficult condition as compared to the difficult-easy condition ($p < .05$). Additionally, during Minute 6, there was a significantly lower percent timeouts in the easy control condition as compared to the easy-difficult condition ($p < .01$). Therefore, there were very few significant differences in the performance data for the percent timeouts measure, and those significant differences were only between easy and difficult portions of the trials (for results of Tukey's HSD tests, see Appendix N).

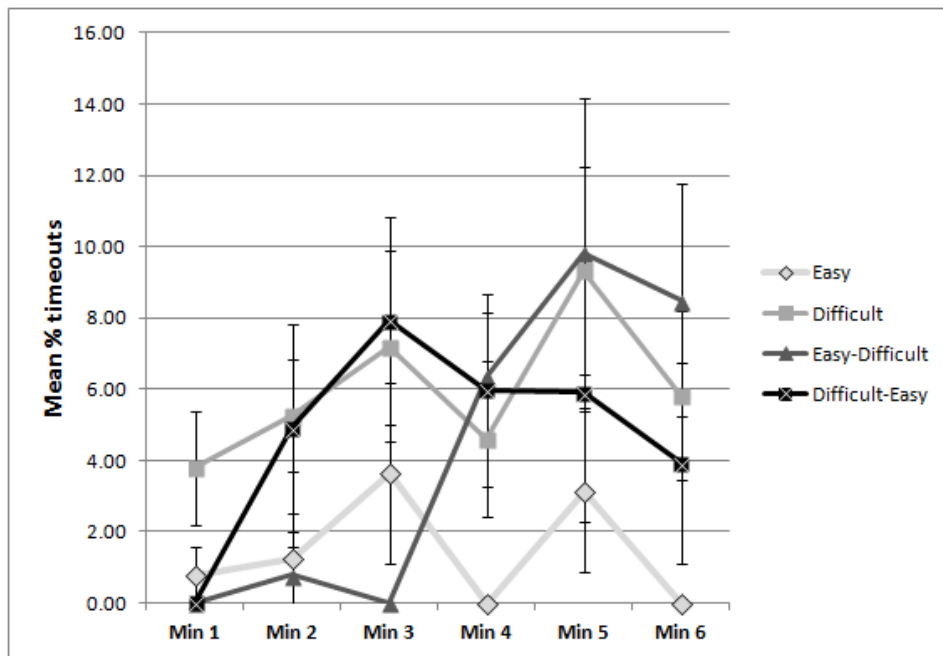


Figure 10. Mean percent timeouts in the Communications task by condition and minute. The error bars represent standard errors.

The one-way ANOVA for mean correct reaction time in the Communications task revealed significant simple main effects for every minute ($F(3, 45) = 6.34$ to 9.54 , $p = .000$ to $.002$, $\eta_p^2 = .044$ to $.696$) except for Minutes 2 and 3.

Figure 11 illustrates mean reaction time of correct responses in the Communications task as a function of difficulty condition and minute. The figure shows that means for each condition are fairly closely clustered for Minutes 1-3. Starting with Minute 4, the fastest

correct mean reaction times are in the difficult-easy condition. Tukey's HSD tests were used to further investigate the significant differences found in the one-way ANOVAs. While significant differences between some of the easy and difficult portions of the trials were found for Minutes 1, 4, 5, and 6, performance in the easy and difficult control conditions was not significantly different for any minute. The failure to find significant differences between the easy and difficult control conditions during any minute makes interpretation of the results problematic (for results of Tukey's HSD tests, see Appendix O).

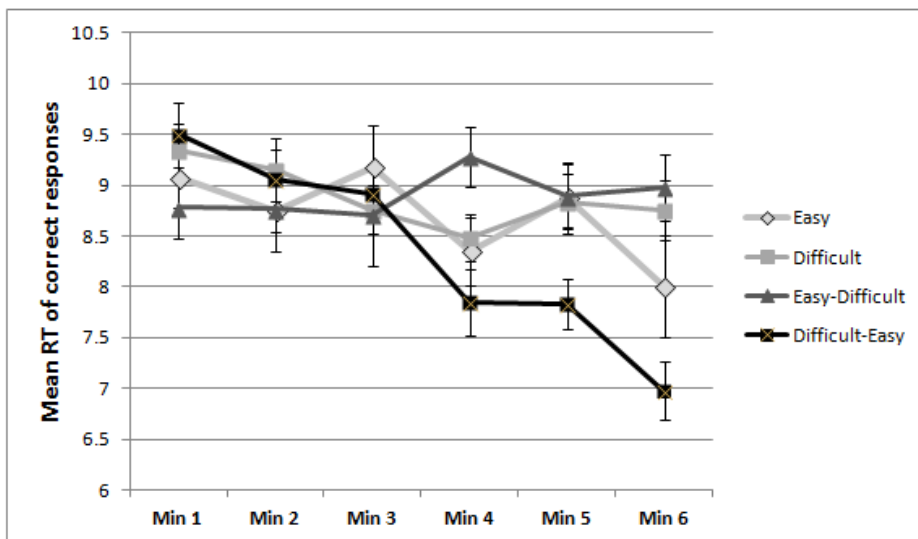


Figure 11. Mean correct reaction time in the Communications task by condition and minute. The error bars represent standard errors.

A closer look at the timeouts and reaction time measures revealed that the Communications task was not ideal for examining workload transitions. There are two factors that contributed to this result. First, there are time constraints associated with the Communications task. It takes time for the instructions in each communication to be spoken, and then participants are allotted 15 seconds to comply. This limits the total number of communications that can be issued in a 6-minute trial. Thus, to ensure differences in the level of difficulty between easy and difficult trials, there were relatively few communications issued during the easy trials. Therefore, the mean reaction time for correct responses in the easy portions of trials was based on only a few responses. This problem was exacerbated by partitioning performance into minute-by-minute increments. The second problem encountered was the ceiling effects that occurred in performance. Regardless of the condition, and as noted previously, participants had very few timeouts. Indeed, the one-way ANOVAs revealed that in a number of the minute-by-minute comparisons with each of these performance measures, there were no significant differences between any of the conditions. As a result, the performance measures from the Communications task (percent timeouts, and the correct mean reaction time) were not considered further in analyses of transition effects.

Conclusions for performance measures

In general, the significant differences between conditions, within each minute, were primarily demonstrated between easy and difficult portions of the trials. However, significant differences were found between the easy and difficult-easy conditions in Minute 4 of the Systems task in percent false alarms and in the Tracking RMS error. The RMS error in the Tracking task and the percent false alarms in the Systems task were both higher in the difficult-easy condition than in the easy control condition during Minute 4. An additional significant difference was demonstrated in the mean correct reaction time between the easy and difficult-easy conditions in Minute 6 of the Systems task, with higher mean correct reaction times occurring in the difficult-easy condition as compared to the easy control condition. These results can be interpreted as lending some support to the position that workload transitions impacted performance in these conditions. However, as described previously, unanticipated trends in performance in the easy condition may have had an impact on the noted differences between the easy and easy-difficult conditions in Minutes 4 and 6 of the Systems task, thereby making clear interpretation of these results somewhat problematic as they pertain to transition effects.

Subjective Measures

The NASA TLX was administered at the end of each trial, and participants were asked to complete it with respect to the workload they experienced during the second half of each trial (i.e., Minutes 4-6). TLX composite scores were calculated by computing the average of the scores for the six sub-scales for each trial (Christ et al., 1993; Hendy, et al., 1993; Nygren, 1991). Then, the composite scores were collapsed across trials within each difficulty condition to create average scores. Figure 12 illustrates the mean, unweighted TLX composite scores as a function of difficulty condition. As can be seen in the figure, the mean composite scores for the easy and difficult-easy conditions are very similar, as are the scores for the difficult and easy-difficult conditions. As illustrated, participants rated the workload in the second half of the easy and difficult-easy trials as lower than in the second half of the difficult and the easy-difficult trials. These results suggest that the participants were successful in rating the workload associated with the second half of the trials within the respective conditions.

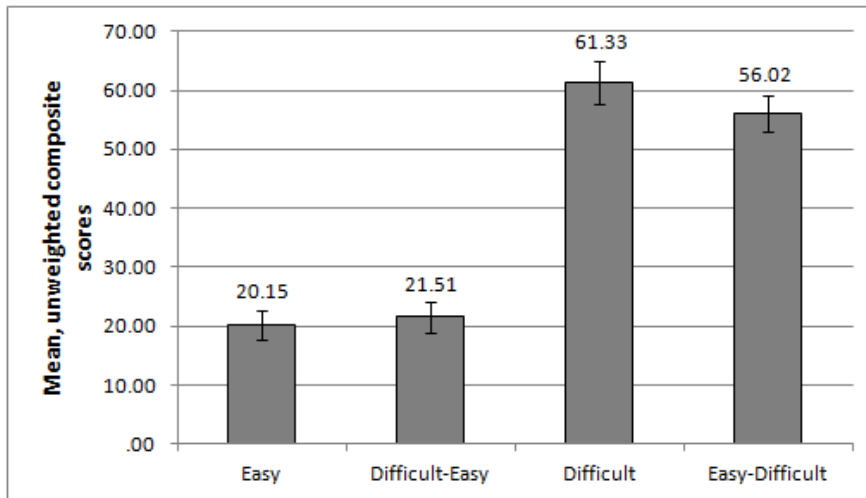


Figure 12. Mean, unweighted composites workload scores for NASA TLX in each condition. Error bars represent standard errors.

A one-way, repeated measures ANOVA was conducted to compare differences between conditions, and it resulted in a significant main effect, $F(3, 45) = 67.61, p < .01, \eta_p^2 = 0.818$. Post hoc Tukey's HSD tests showed that the subjective workload in the easy condition and in the difficult-easy condition was rated as significantly lower than in the easy-difficult condition and difficult conditions ($p < .01$). Because participants were asked to think about the second half of each trial as they completed the TLX, it is not surprising that they gave lower workload ratings to all easy portions of the trials than to the difficult portions of the trials. The comparisons between the easy and difficult-easy conditions and between the difficult and the easy-difficult conditions were of the most interest from the workload transition perspective. However, mean composite workload

scores demonstrated no trends for any substantial differences between these conditions, and analyses confirmed that there were no significant differences that were consistent with workload transition effects.

The unweighted scores from the TLX subscales were also examined. Figure 13 illustrates the mean scores for each subscale as a function of difficulty condition. As shown in the figure, all subscale scores were very similar in the easy and difficult-easy conditions and in the difficult and easy-difficult conditions. Additionally, the scores for the easy and difficult-easy conditions were lower for each subscale than for the comparable scores in the difficult and easy-difficult conditions.

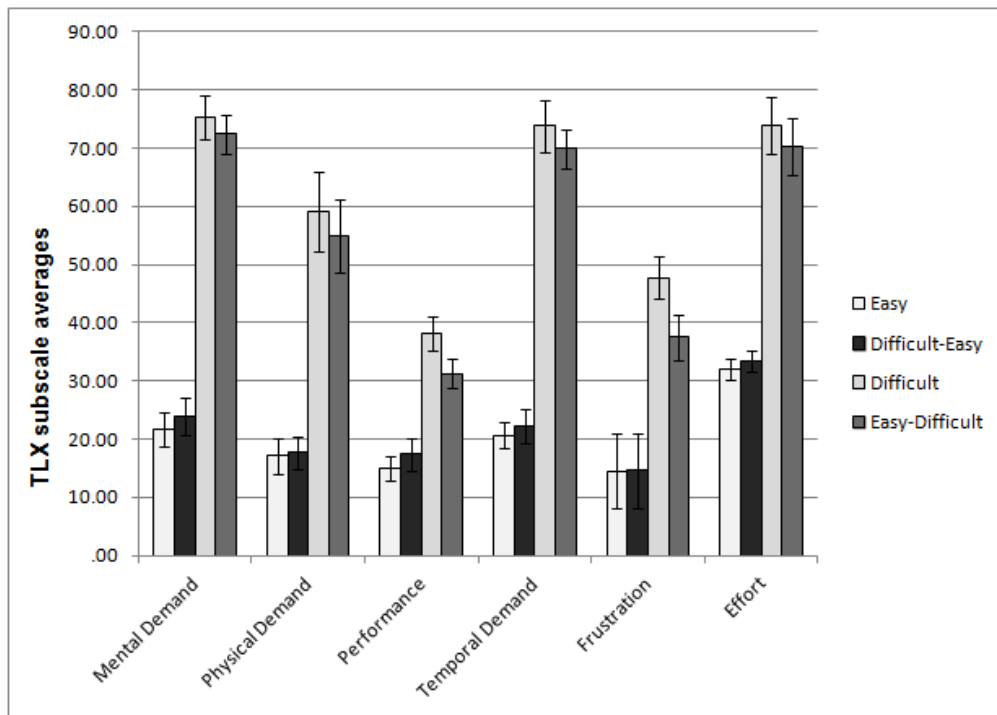


Figure 13. Mean, unweighted composite scores for each TLX subtask in each difficulty condition. Error bars represent standard errors.

A 4 (difficulty condition) x 6 (subscale) repeated measures ANOVA resulted in a significant main effect of difficulty condition ($F(3, 45) = 20.82, p < .01, \eta_p^2 = 0.818$) and a significant main effect of subscale ($F(5, 75) = 67.61, p < .01, \eta_p^2 = 0.581$). There was also a significant interaction between difficulty condition and subscale ($F(15, 225) = 15.38, p < .01, \eta_p^2 = .506$). To further investigate the significant interaction, one-way ANOVAs were conducted for each subscale, and they were all significant for the simple main effect of difficulty condition (for ANOVA results, see Appendix G, Table 4).

Tukey's HSD tests were used to investigate the significant main effects. The results indicated that participants felt that the easy condition involved significantly less mental demand ($p < .01$), physical demand ($p < .01$), temporal demand ($p < .01$), effort ($p < .01$), and frustration ($p < .01$) than the easy-difficult and difficult control conditions. Participants also felt they performed better ($p < .01$) in the easy condition than in the easy-difficult and the difficult conditions. The same pattern of significant differences was found in comparison between the difficult-easy condition and both the difficult and the easy-difficult conditions ($p < .01$). In these comparisons, the difficult-easy condition resulted in lower ratings of mental, physical and temporal demand, as well as lower levels of effort and frustration compared with the difficult and the easy-difficult conditions. Similarly, participants felt that they performed better in the difficult-easy condition as compared with the other two conditions (for results of Tukey's HSD tests, see Appendix P). Therefore, none of the significant differences between conditions indicated that the workload transitions impacted participants' perceptions of overall workload or their perceptions of any of the subscales that make up the TLX.

As indicated previously, the shortened DSSQ measures mood changes associated with task performance using three state measures: engagement, distress, and worry. Also as noted above, the shortened DSSQ was administered prior to AF-MATB performance and then again at the conclusion of each block of trials. One overall score is generated

for each of the three measures. Unlike its original version, however, the shortened DSSQ has no normal sample population with which to compare study scores. Without the normal sample population, comparisons were accomplished by standardizing the shortened DSSQ raw scores. To standardize the scores, the mean and standard deviation of the pre-performance (baseline) scores were first calculated for the group for each state measure. Then, the group's pre-performance mean and standard deviation were used to transform each individual's raw scores for both the pre-performance and post-performance results (raw score – group average/ group SD), which is the method advocated by the creators of the shortened DSSQ. In this way, the group of participants served as its own normal sample population. Additionally, this standardization reduced the variability often seen in response to subjective questionnaires caused by individual differences in response styles. Following the standardization of the scores, difference scores between the pre-performance and post-performance z-scores were calculated. Finally, average difference scores were calculated for each state measure in each condition.

The average difference scores for the distress and engagement measures are illustrated as a function of difficulty condition in Figure 14. The figure shows that engagement decreased from baseline for all of the conditions, and participants reported being least engaged during the easy condition. Also, Figure 14 shows that levels of

distress dropped below baseline levels only in the easy condition. It also appears that, of all of the conditions, the difficult condition had the highest self-reported levels of distress. The last state measure, worry, is shown as a function of condition in Figure 15. Worry appears to increase from baseline in the easy and difficult conditions, and it decreased for the easy-difficult and difficult-easy conditions.

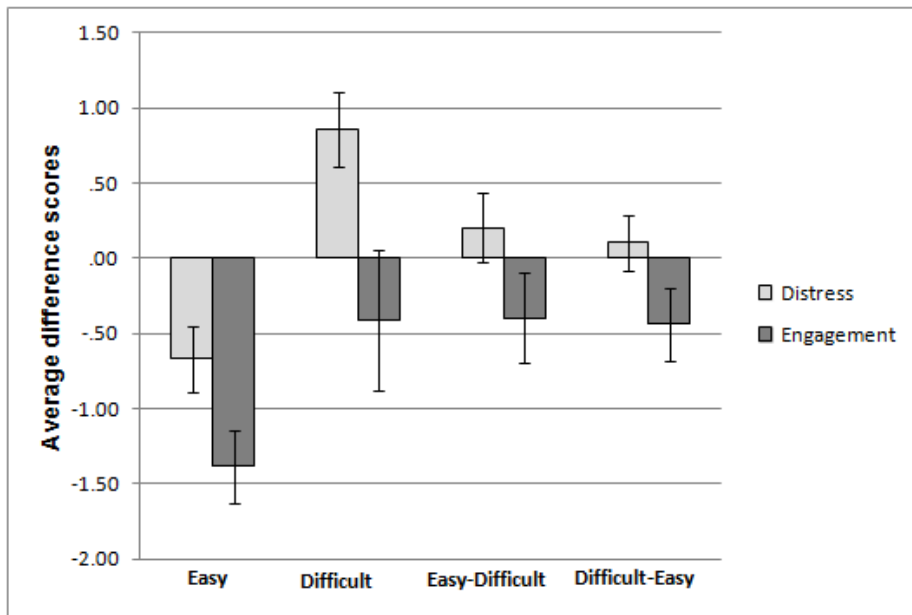


Figure 14. Average differences scores (z-scores) of distress and engagement from the shortened DSSQ as a function of condition. Error bars represent standard errors.

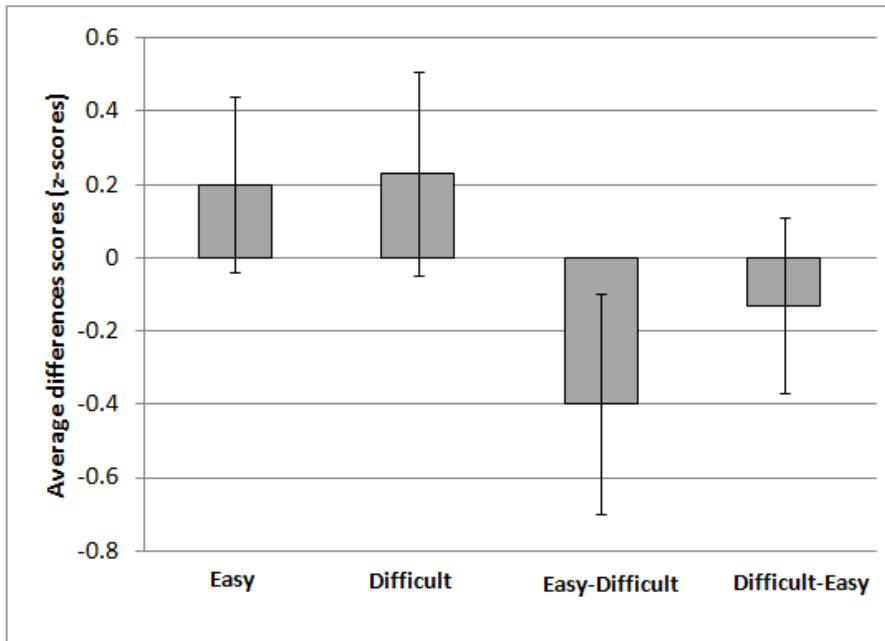


Figure 15. Average differences scores (z-scores) of worry from the shortened DSSQ as a function of condition. Error bars represent standard errors.

The differences scores were then used in one-way, repeated measures ANOVAs to determine the effect of difficulty condition on each of the state measures. These three ANOVAs resulted in a non-significant effect of condition for worry ($F(3, 45) = 1.90, p > .05, \eta_p^2 = 0.284$), but significant effects for both engagement ($F(3, 45) = 8.41, p < .01, \eta_p^2 = 0.359$) and distress ($F(3, 45) = 10.90, p < .01, \eta_p^2 = 0.113$). To follow up, Tukey's HSD tests were conducted. The results demonstrated that the difficult condition, the easy-difficult condition, and the difficult-easy condition were all significantly more

engaging than the easy condition ($p < .01$). The analyses also showed that the difficult control condition was significantly more distressing than the easy condition ($p < .01$).

These results illustrate that the two types of transitions were not perceived as being significantly different from each other in terms of engagement, distress or worry. The results also demonstrate that the blocks of transition trials did not differ significantly from the blocks of control trials in perceived levels of distress or worry. Finally, the results illustrated that all of the blocks of trials were more engaging than the easy block of trials. Engagement also decreased from baseline for all of the conditions, but this is a common result that is often due to participants' growing disinterest in the task as time progresses over the course of the data collection session.

In conclusion, analysis of the TLX unweighted composite scores found that the easy and the difficult-easy conditions were rated as having significantly lower workload than the difficult and easy-difficult conditions. Additionally, examination of the six TLX subscale scores revealed that the easy and difficult-easy conditions resulted in significantly lower ratings of workload on each dimension than the difficult and easy-difficult conditions. The shortened DSSQ showed that participants found that all of the conditions involving a difficult component were more engaging than the easy control condition. It also revealed that distress was significantly higher in the difficult control condition than the in easy control condition and that there were no significant differences

between conditions in terms of worry. Most importantly, from the perspective of the current study, neither the TLX nor the shortened DSSQ results provided any evidence of workload transition effects.

CHAPTER 5

DISCUSSION

Main Findings

This study used a multi-tasking environment and subjective measures to examine the effects of workload transitions on performance and on participants' perceptions of task-related stress and workload. The results of the subjective measures, the NASA TLX and the shortened DSSQ, did not indicate that workload transitions in the AF-MATB environment have any significant positive or negative effect on perceptions of workload levels or on task-related stress. On the other hand, the performance data did yield results indicating that transitioning from a high level of workload to a low level can be detrimental to performance, as compared to baseline performance. Similar performance results, in regards to transition effects, were found in several other studies (Cox-Fuenzalida & Angie, 2005; Cox-Fuenzalida, Beeler & Sohl, 2006; Cox-Fuenzalida, Swikert, & Hittner, 2004; Cumming and Croft, 1973; Goldberg and Stewart, 1980; Farrell, 1999; Gluckman et al., 1993; Matthews, 1986). Although a number of theories

have been posited to account for workload transition effects, the current results are most clearly interpretable within the frameworks associated with the theory of strategic persistence (Matthews, 1986) or the resource depletion theory (Ungar et al., 2005).

Subjective Measures

Analysis of the NASA TLX unweighted composite scores revealed that the only significant differences were between the easy and difficult portions of the trials, with the former resulting in lower ratings of workload than the latter. Likewise, the unweighted subtask scores were significantly lower for the easy and difficult-easy conditions as compared to the difficult and easy-difficult conditions. Thus, participants' perceived workload across both composite and individual subscale measures was lower in easy and difficult-easy conditions than in the difficult and easy-difficult conditions, a result which confirms that the manipulation of task difficulty affected perceived workload. Most importantly from the perspective of the current investigation, none of the results indicated that the transition trials were perceived as having different levels of workload than their respective controls.

Similar results were found in studies by Moroney et al. (1993) and Fischer et al. (1994) that used Microsoft's "Flight Simulator" which is similar to the AF-MATB

tracking task. In these studies, there were no significant differences in the performance scores or in the TLX scores when the transition trials were compared with their corresponding controls. However, a study by Hancock et al. (1995) used a compensatory tracking task and found that participants, using both the TLX and SWAT, rated a medium difficulty condition as having lower workload after just completing a more difficult condition.

The results of shortened DSSQ revealed that there was no significant main effect of difficulty condition for worry. This indicates that the workload transitions did not have a significant impact on levels of worry. On the other hand, there were main effects of difficulty condition for distress and engagement. Engagement decreased in all of the conditions, as compared to pre-performance baseline levels, and analyses also showed that the difficult, easy-difficult, and difficult-easy conditions were all significantly more engaging than the easy condition. With this measure, declines in engagement are not unusual over the course of the data collection session. Additionally, distress fell below pre-performance baseline levels only in the easy condition. The analysis showed that the only significant differences in distress were between the easy and difficult control conditions, with the latter exhibiting higher levels of distress than the former. These results indicate that the transitions in difficulty had little or no impact on participants'

feelings of distress or engagement beyond the differences observed between difficult and easy control conditions.

These findings are inconsistent with those of Helton et al. (2004) who obtained significant workload transition results with the DSSQ, despite not finding significant performance differences. In their signal-salience based vigilance study, Helton et al. found that the transitioned participants experienced more distress than non-transitioned controls. They also found that shifts from high to low signal salience (i.e., low to high workload) reduced participants' task engagement, while shifts from low to high signal salience (i.e., high to low workload) had the opposite effect. In a second study utilizing the same vigilance paradigm, Helton et al. (2008) also reported that workload transitions resulted in higher levels of distress than were experienced by participants in non-transition conditions, despite warnings that transitions might occur. Transitions from either high to low or low to high signal salience also lowered ratings of engagement relative to controls in unwarned conditions. In contrast, the current study found that distress was lowest in the easy condition and highest in the difficult condition. Meanwhile, the easy-difficult and difficult-easy transition conditions had very similar levels of distress and did not differ significantly from either control condition. Additionally, the current study found that levels of engagement in the difficult, easy-

difficult, and difficult-easy conditions were significantly higher than in the easy condition.

One factor contributing to the differences in results between the current study and the Helton et al. experiments might have been Helton et al.'s use of the full DSSQ. The full DSSQ may be more sensitive to transition effects than the shortened version. It is also possible that the task-related stress experienced by participants during a vigilance task is simply different than that experienced during a multiple element task battery like the AF-MATB that incorporates several types of tasks. Differences in both workload and the length of trials between the task conditions used in the respective experiments could also have contributed to the different DSSQ results that were obtained.

Performance Measures

The initial ANOVAs (4 (difficulty condition) x 6 (trial)) indicated that there were few trial-based repetition effects. However, there were significant main effects of trial for Tracking RMS error and for average deviation in the Resource Management task. Performance in the Tracking task consistently declined from the beginning to the end of the study. It is possible that participants experienced some fatigue effects that were manifested in the tracking task. This trend could add some ambiguity to the interpretation of the results of this subtask. However, it was determined unlikely that the

noted trials effect would have a large impact on the analysis of minute by minute performance data, beyond an increase in the variance of each minute. Indeed, the analysis of the minute by minute data did not suggest that fatigue affected performance at this level.

The differences in mean deviation in the Resource Management task appeared to be less systematic. Deviations during Trial 2 were significantly higher than during Trial 3, Trial 4, and Trial 5. This may simply be the product of participants' shifting their task priorities during performance of the multiple tasks within the AF-MATB. It is possible that the participants shifted attention between subtasks in such a way that performance in the Resource Management task suffered during Trial 2. However, beyond an increase in the variance of each minute, it was concluded that this trials effect would not have a large impact on the analysis of the minute by minute performance data.

Next, the data were collapsed across trials and partitioned into minute-by-minute increments. The 4 (condition) x 6 (minutes) repeated measures ANOVAs produced significant main effects of difficulty condition for all of the performance measures and significant main effects of minute for some. There were also significant interactions for all of the performance measures, except for percent correct in the Communications task. To probe these interactions, one-way ANOVAs were then used to compare performance within each minute as a function of condition. Based on these analyses and the patterns

in the data, it was concluded that the performance measures related to the Communications task were not ideal for these comparisons. Typically fewer than six communications were issued during each minute. Therefore, the performance means for each minute were based on very few observations and would be expected to be variable and unstable as a result. Consequently, the percent timeouts and mean correct reaction time in the Communications task were not considered further in the analyses of transition effects.

Many of the one-way ANOVAs conducted on the performance measures subsequent to significant interactions did yield significant differences between conditions within minutes. Tukey's HSD tests were used to follow up these significant simple main effects. For the most part, the significant differences in the Tukey's tests indicated that the easy portions of the trials produced significantly better performance than the difficult portions of trials, across all conditions. However, there were significant differences found for several measures between the easy control condition and the difficult-easy condition, post transition. During Minute 4 of the difficult-easy condition, performance in both the Tracking task and in the percent false alarms in the Systems task was significantly worse than in Minute 4 of the easy control condition. Additionally, performance in the mean correct reaction time in the Systems task was also significantly worse in the difficult-easy condition than in the easy condition during Minute 6. These

results suggest that a shift from high workload to low workload can be detrimental to performance. However, the previously described unanticipated trends in performance within the easy control condition in the Systems task percent false alarms and mean correct reaction time made clear interpretation of these results problematical with respect to transition effects. This is especially true of the performance trends for mean correct reaction time in the easy control condition. Therefore, the clearest evidence of transition effects was demonstrated in Tracking RMS error during Minute 4. In this regard, the continuous nature of the Tracking task may have permitted identification of transition effects that might not have been reflected in less continuous tasks (e.g., Communications and Systems tasks), thereby rendering Tracking more sensitive to transition effects than more discrete tasks. Nevertheless, the continuous Resource Management task failed to yield evidence of any transition effect. However, it can be argued that many participants, in an effort to cope with the demands of the various subtasks, may not have treated the Resource Management task like a continuous task. The tanks in the Resource Management task continuously emptied or filled depending on the pumps the participants activated, but the participants monitored fuel levels and adjusted the active fuel pumps at distinct intervals as part of their performance strategy. Therefore, the participants were not continuously active in the Resource Management task in the same way that they were

in the Tracking task, which may have contributed the lack of evidence of any transition effect in the Resource Management task.

In previous workload transition related research, a number of theories have been put forth to explain the various results obtained. An early theory was that transition effects were the result of expectancy effects (Cumming & Croft, 1973; Goldberg & Stewart, 1973; Farrell, 1999; Krulewitz, Warm & Wohl, 1975). The expectancy effects theory explains that performance declines after a transition because participants do not immediately realize that the shift in workload has occurred. Studies examining this theory cued participants to shifts in workload with visual and auditory indicators. These experiments found that shifts from high to low workload negatively impacted performance, despite the cues (Goldberg & Stewart, 1973; Farrell, 1999). Another theory, the theory of contrast effects, states that the first level of task difficulty that participants experience acts an 'anchor' with which they compare any subsequent changes in workload. Therefore, participants who are transitioned from a high level of workload to a low level would perform better than controls, and the opposite would be true for those transitioned from low workload to high. Gluckman et al. (1993) hypothesized that contrast effects were responsible for the transition effects in performance that were previously found by Krulewitz et al. (1975), but a subsequent vigilance study performed by Gluckman et al. produced opposite results of what would

have been predicted by the theory. A third theory proposed that the transition effects were due to overloads in short-term memory capacities. Matthews (1986) conducted two related studies with a task that had no memory component but still found that workload transitions had a significant impact on performance.

The pattern of results observed in the current study can be interpreted within the frameworks proposed by two additional theories, although the current results are not completely consistent with either position. The first of these positions is the resource depletion theory (Ungar et al., 2005). It states that individuals only have a certain amount of cognitive resources to expend on a given task, and more difficult tasks require the expenditure of more cognitive resources. According to this reasoning, participants' performance would decline after a transition from a high level of workload to a lower level, and performance would be better after a transition from a low level to high level of workload as compared to controls. In the current study, this theory would explain both the higher percent of false alarms seen in the Systems task and the higher RMS error observed in the Tracking task during Minute 4 of the difficult-easy condition as compared to the same minute in the easy control condition. Minutes 5 and 6 of the difficult-easy condition could have permitted participants to reamass the cognitive resources needed to attain performance levels that corresponded with the easy condition, as was seen in the current results. However, this explanation does not successfully

address the higher mean correct reactions times in the Systems task in the difficult-easy condition as compared to the easy condition in Minute 6 of the present experiment. Furthermore, the resource depletion theory suggests that performance in the easy-difficult condition, following the transition, should have been better than in the difficult control condition, but the current study failed to find any significant differences in that direction.

The second, partially applicable, theoretical position is the theory of strategic persistence. As described by Matthews (1986), this position explains that transition-related changes occur in performance because participants fail to switch to the optimum strategy, even though they recognize that the level of demand has changed. Matthews expected that participants who were transitioned from a low level of workload to a high level would maintain their low level of response and would perform worse compared to high workload controls who did not transition. Likewise, participants transitioning from a high level of workload to a lower level would maintain their high level of responding and perform better than low workload controls. In the present study, transitions from a high level of workload to a lower level did not produce better performance than the control condition. However, it is possible that a high level of responding in the difficult portion of the trial could translate into a high level of false alarms in the Systems task and over-corrections in the Tracking task after the transition to low workload. This could explain the higher levels of false alarms and the higher tracking RMS error seen in

Minute 4. Additionally, a significantly slower correct mean reaction time in the Systems task was found for Minute 6 in the difficult-easy condition as compared to the easy condition. It is possible that participants recognized that their rate of false alarms was high in Minute 4, and in effort to improve performance, they began slowing their responses. However, the before mentioned decline in mean reaction time in the easy control condition between Minute 5 and Minute 6 make any transition-related interpretations difficult to support.

The theory of strategic persistence, however, is not fully supported by the current data. According to this theory, performance should have been worse in the easy-difficult condition than in the difficult control condition, and in the current study, there was a lack of significant differences obtained when participants were transitioned from low to high levels of workload as compared to the high workload control condition.

Therefore, the lack of significant differences found between the easy-difficult and difficult conditions prevent these data from fully supporting either theory. One argument that may help explain the asymmetric difference found in this study is based on physiological evidence. The human sympathetic nervous system is very efficient in taking people from a resting state to one in which they are fully equipped to flee or fight off an attack. In fact, even the mobilization of resources for non-threatening activities, like exercise, generally occurs swiftly. In contrast, it can take 10 minutes or more for the

body to fully return to its resting state after such stressors (Borresen & Lambert, 2008). Perhaps the brain's preparation to respond to high cognitive demands occurs as swiftly, which would mean that people are well equipped to adjust from a low workload level to high level without performance decrements. This possibility would explain the absence of transition-related differences when performance in the easy-difficult condition is compared to that of the difficult control condition. Asymmetric differences like these were also found in several other workload transition studies (Cumming and Croft, 1973; Goldberg and Stewart, 1980; Farrell, 1999; Gluckman et al., 1993). Additionally, Fischer, et al. (1994), Moroney, et al. (1993), Helton et al. (2004), and Helton et al. (2008) also found the transitions from high workload to low workload had no significant impact on performance, although transitions in the other direction also failed to produce significant differences relative to the appropriate control in these studies.

Conclusion

With the use of the AF-MATB, the current study sought to discover how workload transitions might affect performance in a multitasking environment and to explore associated subjective experiences through the use of the NASA TLX and shortened DSSQ. The results of the NASA TLX and shortened DSSQ did not provide support for the position that workload transitions affect either subjective assessments of workload or

task-related stress. Perhaps a workload measure with more sensitivity than the NASA TLX may have provided different results. Likewise, potential differences in sensitivity between the DSSQ and the shortened DSSQ, in terms of detecting transition effects, may have influenced the results. Further investigation of the latter may be needed. However, the transition related performance results that were obtained were modest, which may have contributed to the lack of transition-related significant results in the subjective measures.

Additionally, despite the many performance measures provided by the AF-MATB task, most of the significant differences between conditions were between the easy and difficult portions of trials. A major aim of the study was to examine the effects of workload transitions in a multi-tasking environment, but individual differences in participants' approach to factors such as subtask prioritization may have obscured any transition effects that may have been present in performance. Assigning primary-task status to one subtask and rank ordering the importance of the other subtasks could provide an interesting view into workload transitions in real-world situations where some subtasks are more critical than others. However, research pertaining to resource theory that has used primary and secondary tasks (Parasuraman & Hancock, 2001) has shown that there are pitfalls associated with this method as well.

The results directly relating to transition effects were relatively modest, although the data from several measures were consistent with the position that performance was negatively affected, post-transition, when trials began at a high level of workload and then became less demanding. As indicated previously, however, only one of these measures, RMS tracking error, could be unambiguously interpreted as supporting the presence of such a transition effect. Two theoretical frameworks, the resource depletion theory and the theory of strategic persistence were both discussed as possible explanations of these results, although the current data set did not entirely fit either theory. Further research on this topic, particularly in multi-tasking environments, is needed.

REFERENCES

- Alluisi, E.A., Muller, P.F. & Fitts, P.M. (1957). An information analysis of verbal and motor responses in a forced-pace serial task. *Journal of Experimental Psychology*, 53 (3) 153-158.
- Bailey, L., Broach, D., Thompson, R., & Enos, R. (1999). Controller teamwork evaluation and assessment methodology (CTEAM): A scenario calibration study (Dot/FAA/AM-99/24). Washington, DC: Federal Aviation Administration, Office of Aviation Medicine.
- Bakan, P. (1959). Extraversion-introversion and improvement in an auditory vigilance task. *British Journal of Psychology*, 50, 325-332.
- Borresen, J. & Lambert, M.I. (2008). Autonomic control of heart rate during and after exercise: Measurements and implications for monitoring training status. *Sports Medicine*, 38(8), 633-646.
- Chamberlain, G.P. (1968). Human information transmission under varying demand. M. Eng. Sc. thesis, University of Melbourne.

- Christ, R.E., Hill, S.G., Ayers, J.C., Iavecchia, H.M., Zaklad, A.L., & Bittner, A.C. (1993). Application and validation of workload assessment techniques. *Technical Report 974. U.S. Army Research Institute for the Behavioral Sciences, Alexandria, VA.*
- Comstock, J.R., & Arnegard, R.J. (1992). The Multi-Attribute Task battery for human operator workload and strategic behavior research. NASA TM-104174, National Aeronautics and Space Administration, Langley Research Center, 1992.
- Colquhoun, W.P. (1961). The effect of "unwanted" signals on performance in a vigilance task. *Ergonomics*, 4, 41-52.
- Comstock, J.R.J, Arnegard, R.J., (1992). The multi-attribute task battery for human operator workload and strategic behavior research. National Aeronautics and Space Administration technical Memorandum, No. 104174.
- Cox-Fuenzalida, L.E., & Angie, D.A. (2005). The effect of workload history on dual task performance. *Current Psychology*, 24(3), 171-179.
- Cox-Fuenzalida, L.E., Beeler, C., & Sohl, L. (2006). Workload history effects: A comparison of sudden increases and decreases on performance. *Current Psychology*, 25(1), 8-14.

- Cox-Fuenzalida, L.E., Swickert, R., & Hittner, J.B. (2004). Effects of neuroticism and workload history on performance. *Personality and Individual Differences*, 36, 447-456.
- Cumming, R.W., & Croft, P.G. (1973). Human information processing under varying task demand. *Ergonomics*, 16(5), 581-586.
- Eysenck, H.J. (1967). *The biological basis of personality*. Springfield, IL: Charles C. Thomas
- Farrell, P.S.E. (1999). The hysteresis effect. *Human Factors*, 41(2), 226-240.
- Fischer, D.S., Moroney, W.F., & Beiers, D.W. (1994). Workload context effect: An elusive phenomenon. Proceedings of the Human Factors and Ergonomics Society 38th Annual meeting.
- Gibbs, C.B. (1965). Probability learning in step-input tracking. *British Journal of Psychology*, 56, 233-242.
- Gluckman, J.P., Warm, J.S., Dember, W.N., Rosa, R.R. (1993). Demand transitions and sustained attention. *The Journal of General Psychology*, 120(3), 323-337.
- Goldberg, R.A., & Stewart, M.R. (1980). Memory overload or expectancy effect? 'Hysteresis' revisited. *Ergonomics*, 23(12), 1173-1178.

- Hancock, P.A., Williams, G., Manning, C.M., & Miyake, S. (1995). Influence of task demand characteristics on workload and performance. *The International Journal of Aviation Psychology*, 5(1), 63-86.
- Hancock, P.A., Williams, G., Miyake, S., & Manning, C.M. (1992, unpublished). The influences of task demand characteristics on workload and performance. University of Minnesota.
- Hart, S.G., & Staveland, L.E. (1988). Development of a multi-dimensional workload scale: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.
- Helton, W.S., Shaw, T.H., Warm, J.S., Matthews, G., Dember, W.N., & Hancock, P.A. (2004). Demand transitions in vigilance: Effects on performance efficiency and stress. In: D.A. Vincenzi, M. Mouloua., and P.A. Hancock (Eds.). *Human Performance, situation awareness and automation: Current research and trends.* (pp. 258-262), Erlbaum: Mahwah, NJ.
- Helton, W.S., Shaw, T.H., Warm, J.S., Matthews, G., & Hancock, P.A. (2008). Effects of warned and unwarned demand transitions on vigilance performance and stress. *Anxiety, Stress, & Coping*, 21 (2), 173-184.

- Hendy, K.C., Hamilton, K.M., & Landry, L.N. (1993). Measuring subjective workload: When is one scale better than many? *Human Factors*, 35, 579-601.
- Huey, B.M., & Wickens, C.I. (Eds.). (1993). Workload transition: Implications for individual and team performance. Washington DC: National Academy Press. (National Research Council).
- Kruewitz, J., Warm, J.S., & Wohl, T.H. (1975). Effects of shifts in the rate of repetitive stimulation on sustained attention. *Perception & Psychophysics*, 18(4), 245-249.
- Luximon, A., & Goonetilleke, R.S. (2001). Improvement of the Subjective Workload Analysis Technique. *Ergonomics*, 44, 229-243.
- Mackworth, J.F. (1968). Vigilance, arousal, and habituation. *Psychological Review*, 75, 308-322.
- Mackworth, J.F. (1970a). *Vigilance and attention*. Baltimore: Penguin.
- Mackworth, J.F. (1970b). *Vigilance and habituation*. Baltimore: Penguin.
- Matthews, M.L. (1986). The influence of visual workload history on visual performance. *Human Factors*, 28(6), 623-632.
- Matthews, G., Campbell, S.E., Falconer, S., Joyner, L.A., Huggins, J., Gilliland, K., Grier, R., & Warm, J.S. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress and worry. *Emotion*, 2, 315-340.

- Matthews, G., Emo, A.K., & Funke, G.J. (2005). A short version of the Dundee Stress State Questionnaire. *Presented at the Twelfth Meeting of the International Society for the Study of Individual Differences*, Adelaide, Australia, July 2005.
- Matthews, G., Joyner, L., Gilliland, K., Huggins, J., & Falconer, S. (1999). Validation of comprehensive stress state questionnaire: Towards a state big three? In I. Merville, I.J. Deary, F. DeFruyt, & F. Ostendorf (Eds), *Personality psychology in Europe* (Vol. 7, pp. 335-350). Tilburg: Tilburg University Press.
- Microsoft (1990). *Flight simulator: Aircraft and scenery designer*. Washington: Microsoft.
- Microsoft (1990). *Flight simulator*. Washington: Microsoft.
- Miller, W. D. (2010). The U.S. Air Force-developed adaptation of the Multi-Attribute Task Battery for the assessment of human operator workload and strategic behavior (Technical Report AFRL-RH-WP-TR-2010-0133). Retrieved from the Defense Technical Information Center website:
<http://www.dtic.mil/dtic/tr/fulltext/u2/a537547.pdf>
- Morgan, J.F., & Hancock, P.A. (2011). The effect of prior task loading on mental workload: An example of hysteresis in driving. *Human Factors*, 53(1), 75-86.

- Moroney, B.W., Warm, J.S., & Dember, W.N. (1995). Effects of demand transitions on vigilance performance and perceived workload. Proceedings of the Human Factors and Ergonomics Society 39th annual meeting.
- Moroney, W.F., Reising, J., Biers, D.W., & Eggemeier, F.T. (1993). The effect of previous level of workload on the NASA Task Load Index (TLX) in a simulated flight task. Proceeding of the 7th International Symposium on Aviation Psychology, 882-885.
- Nygren, T.E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors*, 35, 579-601.
- Parasuraman, R., & Hancock, P.A. (2001). Adaptive control of mental workload. In P.A. Hancock & P.A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 305-320). Mahwah, NJ: Erlbaum.
- Poulton, E.C. (1982). Influential companions: Effects of one strategy on another in the within-subjects designs of cognitive psychology. *Psychological Bulletin*, 91, 673-690.
- Reid, G.B. & Nygren, T.E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P.A. Hancock & N. Meshkati, (Eds.), *Human mental workload* (pp. 185-214). Amsterdam: Elsevier.

Sternberg, S. (1969). On the discovery of processing stages: Some extensions of Donders' method. *Acta Psychologica*, 30, 315-318.

Tabachnick, B. L., & Fidell, L. S. (2007). *Using multivariate statistics (5th ed.)*. Boston, MA: Pearson Education, Inc.

Tukey, J.W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Ungar, N.R., Matthews, G., Warm, J.S., Dember, W.N., Thomas, J.K., Finomore, V.S., & Shaw, T.H. (2005). Demand transitions and tracking performance efficiency: Structural and strategic models. Proceedings of the Human Factors and Ergonomics Society 49th annual meeting.

APPENDIX A:
Material for Recruitment

**Participants Needed for a
Research Project at the
Air Force Research
Laboratory**



The U.S. Air Force Research Laboratory (AFRL) is conducting a study in which participants will be tested on their ability to perform a computerized aviation simulation test. This study is part of an effort to help improve the understanding of how changes in the difficulty of a task may affect an operator's performance on that task. The results of this testing may provide the military and scientific communities with information that could change the operational task environment for operators of remotely piloted aircraft. This study can contribute to efforts related to other mission performance and interface design studies that may aid operators in maintaining optimal performance. Volunteers will be asked to come to the lab for extensive training using the aviation simulation test known as the Multi-Attribute Task Battery (MATB). The MATB requires participants to perform a tracking task while concurrently monitoring warning lights and dials, responding to computer-generated auditory requests to adjust radio frequencies, and managing simulated fuel flow rates using various key presses. This task is accomplished using a personal computer equipped with a standard keyboard, joystick, and mouse. During the testing session, physiological signals, including electroencephalography, cardiac measures, and eye movement will be

recorded. Performance and subjective data will also be collected. The testing session will include several simulations that vary in difficulty and will last approximately four hours. Participants will be compensated for their time. This research project has been approved for the use of human subjects by the Air Force Research Laboratory's Institutional Review Board in accordance with AFI 40-402 and AFRLI 40-402.

Requirements to participate:

- Must be a U.S. citizen, between 18 and 30 years of age.
- Must have 20/20 vision or corrected vision (glasses/contacts OK), normal color vision, and normal hearing.
- To complete the study, participants must be available for approximately 10 hours of training conducted over several days, and for one day of testing requiring approximately six hours.

For more information please contact Iris Davis (Ball Aerospace & Technologies Corp.):

937-xxx-xxx

Please include 1) your name, 2) a phone number where you can be reached, and 3) a list of the days and times you are available to participate.

DISTRIBUTION A. Approved for Public Release. Distribution Unlimited (88ABW-2012-1522)

APPENDIX B:

NASA-Task Load Index



NASA - Task Load Index

RATING SCALE INSTRUCTIONS

We are not only interested in assessing your performance but also the experiences you had during the different experimental scenarios. Here is a description of the technique that will be used to examine your experiences. This technique was developed by NASA to examine "workload." Workload is a difficult concept to define precisely, but a simple one to understand generally. The factors that influence your experience of workload may come from the task itself, your feelings about your own performance, how much effort you put in, the stress and frustration you felt, and so forth.

Since workload is something that is experienced individually by each person, there are no effective "rulers" that can be used to estimate the workload of different activities. One way to find out about workload is to ask people to describe the feelings they experienced. Because workload can be affected by many different factors, we would like you to evaluate several workload components rather than just provide an overall global estimate. There are six rating scales that you will be using to assess workload. The scale titles and descriptions are listed on a separate page. Please read through the descriptions carefully before using the scale.

To do the ratings of a task, you will be given a blank rating form with all six scales listed and a response area for each scale. You will indicate your rating by placing an "X" in the box on the scale that corresponds best to your experience. (Note that the "Performance" scale has "Good" performance on the left, and "Poor" performance on the right. This has been confusing to some people.)

Your ratings of the different task scenarios will be an important part of the data of this evaluation, so please consider your ratings carefully.

If anything is unclear about the rating process or if you have any questions, please ask the experimenter about them.



NASA - Task Load Index

MENTAL DEMAND [Low / High]

How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?

PHYSICAL DEMAND [Low / High]

How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

TEMPORAL DEMAND [Low / High]

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?

PERFORMANCE [Good / Poor]

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 in accomplishing these goals?

EFFORT [Low / High]

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

FRUSTRATION [Low / High]

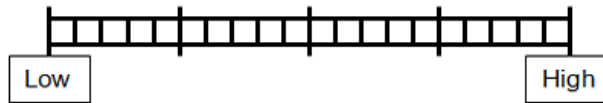
How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

NASA - Task Load Index

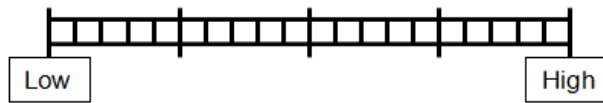
Participant ID: _____

Trial: _____

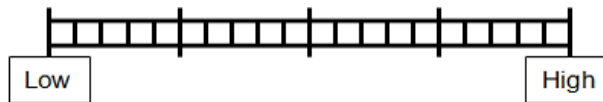
MENTAL DEMAND



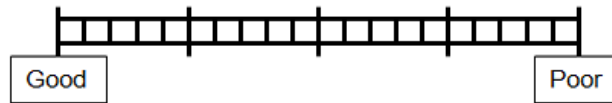
PHYSICAL DEMAND



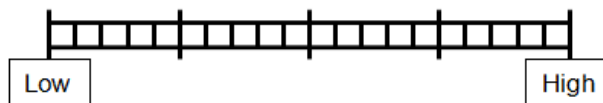
TEMPORAL DEMAND



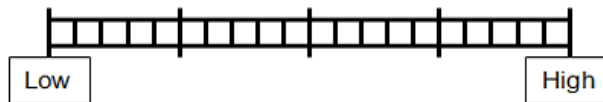
PERFORMANCE



EFFORT



FRUSTRATION





NASA - Task Load Index

SOURCES-OF-WORKLOAD SURVEY INSTRUCTIONS

Several times during the course of the experiment, you rated the task scenarios' workload using the six scales of the NASA – Task Load Index. Scales of this sort are extremely useful, but their utility can suffer from the tendency people have to interpret the scales in individual ways. For example, some people feel that time pressure is the main cause of workload regardless of the level of mental or physical demand. Others experience workload primarily in terms of effort expended. Some people mainly experience workload when some aspect of the task frustrates them. And so forth. Previous studies of workload have discovered a multitude of patterns in the way people use the scales to rate their experiences. And, of course, the tasks being rated can influence how the scales are used as well.

The evaluation that you are about to perform was developed by NASA researchers to assess the relative importance of the six scales in determining the workload you experienced in the experimental tasks.

The procedure is simple. You will be presented with a pair of scale titles on each page of the evaluation form (for example, EFFORT versus MENTAL DEMAND) and asked to choose which was more important to your experience of workload in the tasks that you just performed. Simply circle the scale that you think was more important in your ratings of the tasks.

Do not think that there is any particular correct pattern, this is just a technique to aid in understanding how best to interpret your ratings. The important thing is that the choices accurately reflect your opinion.

If you have any questions, please ask the experimenter.

NASA – TLX Sources of Workload Survey

(Each pair of scale titles should appear on its own page)

Effort or Performance

Temporal Demand or Frustration

Temporal Demand or Effort

Physical demand or Frustration

Performance or Frustration

Physical Demand or Temporal Demand

Physical Demand or Performance

Temporal Demand or Mental Demand

Frustration or Effort

Performance or Mental Demand

Performance or Temporal Demand

Mental Demand or Effort

Mental demand or Physical Demand

Effort or Physical Demand

Frustration or Mental Demand

2. I feel relaxed.	0	1	2	3	4
3. The content of the task will be dull.	0	1	2	3	4
4. I am thinking about how other people might judge my performance.	0	1	2	3	4
5. I am determined to succeed on the task.	0	1	2	3	4
6. I feel tense.	0	1	2	3	4
7. I am worried about what other people think of me.	0	1	2	3	4
8. I am thinking about how I would feel if I were told how I performed	0	1	2	3	4
9. Generally, I feel in control of things.	0	1	2	3	4
10. I am reflecting about myself.	0	1	2	3	4
11. My attention will be directed towards the task.	0	1	2	3	4
12. I am thinking deeply about myself.	0	1	2	3	4
13. I feel energetic.	0	1	2	3	4
14. I am thinking about things that happened to me in the past	0	1	2	3	4
15. I am thinking about how other people might perform on this task	0	1	2	3	4
16. I am thinking about something that happened earlier today.	0	1	2	3	4
17. I expect that the task will be too difficult for me.	0	1	2	3	4
18. I will find it difficult to keep my concentration on the task.	0	1	2	3	4
19. I am thinking about personal concerns and interests.	0	1	2	3	4
20. I feel confident about my performance.	0	1	2	3	4
21. I am examining my motives.	0	1	2	3	4
22. I can handle any difficulties I may encounter	0	1	2	3	4
23. I am thinking about how I have dealt with similar tasks in the past	0	1	2	3	4
24. I am reflecting on my reasons for doing the task	0	1	2	3	4
25. I am motivated to try difficult at the task.	0	1	2	3	4
26. I am thinking about things important to me.	0	1	2	3	4
27. I feel uneasy.	0	1	2	3	4
28. I feel tired.	0	1	2	3	4
29. I feel that I cannot deal with the situation effectively.	0	1	2	3	4
30. I feel bored.	0	1	2	3	4

DSSQ-3 STATE QUESTIONNAIRE

POST-TASK QUESTIONNAIRE

Instructions. This questionnaire is concerned with your feelings and thoughts while you were performing the task. Please answer **every** question, even if you find it difficult. Answer, as honestly as you can, what is true of **you**. Please do not choose a reply just because it seems like the 'right thing to say'. Your answers will be kept entirely confidential. Also, be sure to answer according to how you felt **WHILE PERFORMING THE TASK**. Don't just put down how you usually feel. You should try and work quite quickly: there is no need to think very difficult about the answers. The first answer you think of is usually the best.

For each statement, circle an answer from 0 to 4, so as to indicate how accurately it describes your feelings **WHILE PERFORMING THE TASK**.

Definitely false = 0, Somewhat false = 1,

Neither true nor false = 2, Somewhat true = 3, Definitely true = 4

- | | | | | | |
|--|---|---|---|---|---|
| 1. I felt concerned about the impression I am making. | 0 | 1 | 2 | 3 | 4 |
| 2. I felt relaxed. | 0 | 1 | 2 | 3 | 4 |
| 3. The content of the task was dull. | 0 | 1 | 2 | 3 | 4 |
| 4. I thought about how other people might judge my performance | 0 | 1 | 2 | 3 | 4 |
| 5. I was determined to succeed on the task. | 0 | 1 | 2 | 3 | 4 |
| 6. I felt tense. | 0 | 1 | 2 | 3 | 4 |
| 7. I was worried about what other people think of me. | 0 | 1 | 2 | 3 | 4 |
| 8. I thought about how I would felt if I were told how I performed | 0 | 1 | 2 | 3 | 4 |
| 9. Generally, I felt in control of things. | 0 | 1 | 2 | 3 | 4 |
| 10. I reflected about myself. | 0 | 1 | 2 | 3 | 4 |
| 11. My attention was directed towards the task. | 0 | 1 | 2 | 3 | 4 |
| 12. I thought deeply about myself. | 0 | 1 | 2 | 3 | 4 |
| 13. I felt energetic. | 0 | 1 | 2 | 3 | 4 |
| 14. I thought about things that happened to me in the past | 0 | 1 | 2 | 3 | 4 |
| 15. I thought about how other people might perform on this task | 0 | 1 | 2 | 3 | 4 |
| 16. I thought about something that happened earlier today. | 0 | 1 | 2 | 3 | 4 |

17. I found the task was too difficult for me.	0	1	2	3	4
18. I found it difficult to keep my concentration on the task.	0	1	2	3	4
19. I thought about personal concerns and interests.	0	1	2	3	4
20. I felt confident about my performance.	0	1	2	3	4
21. I examined my motives.	0	1	2	3	4
22. I felt like I could handle any difficulties I encountered	0	1	2	3	4
23. I thought about how I have dealt with similar tasks in the past	0	1	2	3	4
24. I reflected on my reasons for doing the task	0	1	2	3	4
25. I was motivated to try difficult at the task.	0	1	2	3	4
26. I thought about things important to me.	0	1	2	3	4
27. I felt uneasy.	0	1	2	3	4
28. I felt tired.	0	1	2	3	4
29. I felt that I could not deal with the situation effectively.	0	1	2	3	4
30. I felt bored.	0	1	2	3	4

**APPENDIX D:
Consent forms**

INFORMATION PROTECTED BY THE PRIVACY ACT OF 1974

**Informed Consent Document
For
Operator Functional State Assessment in Complex Systems**

711 HPW/RHCP, OASIS Laboratory, Building 840, Wright-Patterson AFB, OH

Principal Investigator/Program Manager: James Christensen, DR-II, 937-938-xxxx,
711 HPW/RHCP

Associate Investigators: Justin Estepp, DR-I, 937-938-xxxx, 711 HPW/RHCP

Krystal M. Thomas, DR-II, 937-938-xxxx, 711 HPW/RHCP

Tyron Gray, 2nd Lt, USAF
AFMC, 711 HPW/RHCP, 937-938-xxxx

Iris Davis, CTR, Human Factors Engineer,
Ball Aerospace, 937-938-xxxx,

Margaret Bowers, CTR, Human Factors Engineer
Ball Aerospace, 937-938-xxxx

Jason Monnin, CTR, Biomedical Engineer
Ball Aerospace, 937-938-xxxx

Melissa Jones, Consortium Research Fellow, Biomedical
Engineer 937-938-xxxx

1. **Nature and Purpose:** You have been asked to volunteer to act as a subject in the research project named above.

The purpose is to determine the workload associated with performing an Uninhabited Air Vehicle (UAV) simulation or the Multi-Attribute Task Battery (MATB), a task that approximates operating a single aircraft or UAV. Your data will be used to determine the difficulty of various scenarios and/or to test methods of implementing task automation. Training on the UAV task will be carried out over several weeks and may require up to 20 hours. Training on the MATB task may require up to 10 hours and will be conducted over several days. Actual data collection sessions will be approximately four hours long. You may be asked to perform the task up to ten times. The research will be conducted at Wright-Patterson Air Force Base, Area B, Building 840, room E200K. There will be approximately 60 subjects in this research effort.

2. **Experimental Procedures:**

In this study, performance and/or physiological data will be collected as you perform either the UAV or the MATB task at several levels of task difficulty. The UAV task involves monitoring the progress of four or more aircraft on a computer screen. During the simulation you will have to watch for threats to the aircraft, change course of the aircraft and send other messages to the aircraft to accomplish the mission of destroying enemy targets. The MATB task requires an operator to perform an unstable tracking task using a joystick while simultaneously monitoring warning lights and dials, responding to simulation-generated auditory requests to adjust radio frequencies, and managing simulated fuel flow rates using various key presses. While performing either task, several physiological measures may be collected. They may include several of the following. Your brain's electrical activity (EEG), eye activity (EOG), and heart rate (ECG) will be recorded. Electrodes will be attached to your head in order to record your brain's electrical activity and your eye movements. Additional electrodes will be attached to your collar bone and chest to record the electrical activity of your heart. Further, your eye activity will be recorded with an

off head eye tracking system. These physiological signals will be used to derive an estimate of your mental workload.

When performing the UAV task, you will sit in front of computer monitors and use a mouse to manipulate a cursor on the screens. When performing the MATB task, you will sit in front of a computer monitor and use a mouse, keyboard and joystick to manage the various task elements. Your performance will be recorded for analysis.

3. **Discomfort and Risks:** The experimental procedures do not present any unusual or risky procedures or equipment. However, you may experience mild skin irritation where the EEG, EOG, and ECG electrodes are placed. Participants with a history of urticaria (hives) should be excluded from the study, as an allergic reaction could occur at the site of electrode placement. If you have fainted or felt lightheaded following a vaccination or blood draw, you will be excluded from participation since placement of the electrodes could cause you to pass out or feel lightheaded. No drugs or medical procedures are involved in this experiment. You will have voice contact with the experimenter at all times. If at any time you feel fatigued, bored, or under a strain, you can inform the experimenter, and a break will be scheduled immediately. Data collected in this study will be treated so as to protect your privacy. Data presented or published will not identify individual participants. Results of this study will be available to you upon request.
4. **Precautions for Female Subjects:** None.
5. **Benefits:** There are no direct benefits from participating in this research. If you are active duty military or a Department of Defense employee, you will receive your normal duty pay. If you are a subject pool member, you will be paid \$15 per hour.
6. **Alternative:** Choosing not to participate is an alternative to volunteering for this study.
7. **Entitlements and Confidentiality:**

- a. Records of your participation in this study may only be disclosed according to federal law, including the Federal Privacy Act, 5 U.S.C. 552a, and its implementing regulations. Your personal information will be stored in a locked cabinet in an office that is locked when not occupied. Electronic files containing your personal information will be password protected and stored only on a DoD server. It is intended that the only people having access to your information will be the researchers named above and the AFRL Wright Site IRB or any other IRB involved in the review and approval of this protocol. When no longer needed for research purposes your information will be destroyed in a secure manner (shredding). Complete confidentiality for military personnel cannot be promised because information bearing on your health may be required to be reported to appropriate medical or command authorities.

- b. Your entitlements to medical and dental care and/or compensation in the event of injury are governed by federal laws and regulations, and that if you desire further information you may contact the base legal office (88 ABW/JA, 257-6142 for Wright-Patterson AFB). In the event of a research related injury, you may contact the AFRL IRB office at 937-656-xxxx.

- c. If an unanticipated event (medical misadventure) occurs during your participation in this study, you will be informed. If you are not competent at the time to understand the nature of the event, such information will be brought to the attention of your next of kin.

Next of kin or designated health care agent (if needed):

Name _____

Phone# _____

- d. The decision to participate in this research is completely voluntary on your part. No one may coerce or intimidate you into participating in this program. You are participating because you want to. James Christensen, 711 HPW/RHCP, or an

associate, has adequately answered any and all questions you have about this study, your participation, and the procedures involved. James Christensen can be reached at (937) 938-xxxx. James Christensen or an associate will be available to answer any questions concerning procedures throughout this study. If significant new findings develop during the course of this research, which may relate to your decision to continue participation, you will be informed. You may withdraw this consent at any time and discontinue further participation in this study without prejudice to your entitlements. The investigator or medical monitor of this study may terminate your participation in this study if she or he feels this to be in your best interest. If you have any questions or concerns about your participation in this study or your rights as a research subject, please contact Col William Butler at 937-656-xxxx.

- e. Your participation in this study may be photographed, filmed and/or audio/videotaped. You consent to the use of these media for training and data collection purposes. Any release of records of your participation in this study may only be disclosed according to federal law, including the Federal Privacy Act, 55 U.S.C. 552a, and its implementing regulations. This means personal information will not be released to unauthorized source without your permission. These recordings will not be used for presentation or publication. They will be stored in a locked cabinet in a room that is locked when not occupied. Only the investigators of this study will have access to these media. They will be maintained for 10 years.

YOU ARE MAKING A DECISION WHETHER OR NOT TO PARTICIPATE.

YOUR SIGNATURE INDICATES THAT YOU HAVE DECIDED TO

PARTICIPATE HAVING READ THE INFORMATION PROVIDED ABOVE.

Volunteer

Signature _____ **Date** _____

Volunteer Name (printed) _____

Advising Investigator Signature _____ **Date** _____

Investigator Name (printed) _____

Witness Signature _____ **Date** _____

Witness Name (printed) _____

We may wish to present some of the video/audio recordings from this study at scientific conventions or use photographs in journal publications. If you consent to the use of your image for publication or presentation in a scientific or academic setting, please sign below.

Volunteer

Signature _____ **Date** _____

Privacy Act Statement

Authority: We are requesting disclosure of personal information. Researchers are authorized to collect personal information on research subjects under The Privacy Act-5 USC 552a, 10 USC 55, 10 USC 8013, 32 CFR 219, 45 CFR Part 46, and EO 9397, November 1943.

Purpose: It is possible that latent risks or injuries inherent in this experiment will not be discovered until sometime in the future. The purpose of collecting this information is to aid researchers in locating you at a future date if further disclosures are appropriate.

Routine Uses: Information may be furnished to Federal, State and local agencies for any uses published by the Air Force in the Federal Register, 52 FR 16431, to include, furtherance of the research involved with this study and to provide medical care.

Disclosure: Disclosure of the requested information is voluntary. No adverse action whatsoever will be taken against you, and no privilege will be denied you based on the fact you do not disclose this information. However, your participation in this study may be impacted by a refusal to provide this information.

APPENDIX E:

Participant Information Form

Participant ID: _____ Date: _____

Sex: M F Age: _____

Are you a US citizen? Yes _____ No _____

Is your first language English? Yes _____ No _____

Are you U.S. active military? Yes _____ No _____

Are you a U.S. Government Civilian Employee? Yes _____ No _____

Vision:

Normal: _____ Corrected: _____ Glasses: _____ Contacts _____

Color Vision: Normal _____ Abnormal _____

Hearing normal? Yes _____ No _____

Handedness: Right: _____ Left: _____ Ambidextrous: _____

Preferred hand for writing: Right: _____ Left: _____

Preferred mouse control: Right: _____ Left: _____

Current Occupation: _____

If Student, enter major: _____

Computer Experience:

Extensive _____

Use Often _____

Moderate Use _____

Light Use _____

Almost Never Use _____

Currently taking Medication? Yes: _____ No: _____

If yes, name of drugs: _____

Allergies? _____

APPENDIX F:

Means and standard errors

F1. Means and standard errors for all performance measures as a function of condition and trial

Condition/Trial	Systems, % correct		Systems, % False Alarms	
	Mean	SEM	Mean	SEM
Easy, Trial 1	94.59	.97	6.26	.80
Easy, Trial 2	93.87	1.40	7.23	1.16
Easy, Trial 3	93.87	1.23	8.50	1.05
Easy, Trial 4	93.87	1.26	6.18	.74
Easy, Trial 5	90.99	1.92	8.70	1.72
Easy, Trial 6	93.27	1.39	6.93	1.15
Hard, Trial 1	82.58	1.94	12.95	1.57
Hard, Trial 2	79.14	2.46	14.15	1.99
Hard, Trial 3	81.39	2.43	13.46	1.62
Hard, Trial 4	79.61	2.78	12.21	1.64
Hard, Trial 5	79.75	2.66	12.93	1.31
Hard, Trial 6	80.82	2.44	13.04	1.37
Easy-Hard, Trial 1	82.74	2.29	10.61	1.36
Easy-Hard, Trial 2	83.78	2.08	10.34	1.68
Easy-Hard, Trial 3	82.00	2.88	11.03	1.75
Easy-Hard, Trial 4	82.91	2.26	11.56	1.39
Easy-Hard, Trial 5	81.82	2.27	11.67	1.76
Easy-Hard, Trial 6	81.95	2.47	11.98	1.88
Hard-Easy, Trial 1	81.74	1.74	12.53	1.88
Hard-Easy, Trial 2	81.50	2.43	11.97	1.98
Hard-Easy, Trial 3	81.12	2.59	12.85	2.02
Hard-Easy, Trial 4	81.72	2.40	11.85	1.86
Hard-Easy, Trial 5	82.16	2.26	11.14	1.28
Hard-Easy, Trial 6	82.29	2.49	11.73	1.56

Condition/Trial	Systems, % timeouts		Systems, correct Mean RT	
	Mean	SEM	Mean	SEM
Easy, Trial 1	5.41	.97	1.55	0.05
Easy, Trial 2	6.13	1.40	1.65	0.04
Easy, Trial 3	6.13	1.23	1.60	0.04
Easy, Trial 4	6.13	1.26	1.68	0.05
Easy, Trial 5	9.01	1.92	1.60	0.05
Easy, Trial 6	6.73	1.39	1.65	0.05
Hard, Trial 1	17.42	1.94	2.01	0.04
Hard, Trial 2	20.86	2.46	1.98	0.04
Hard, Trial 3	18.61	2.43	1.97	0.03
Hard, Trial 4	20.39	2.78	1.95	0.04
Hard, Trial 5	20.25	2.66	2.00	0.04
Hard, Trial 6	19.18	2.44	1.99	0.04
Easy-Hard, Trial 1	17.26	2.29	1.87	0.03
Easy-Hard, Trial 2	16.22	2.08	1.93	0.05
Easy-Hard, Trial 3	18.00	2.88	1.88	0.04
Easy-Hard, Trial 4	17.09	2.26	1.83	0.03
Easy-Hard, Trial 5	18.18	2.27	1.85	0.04
Easy-Hard, Trial 6	18.05	2.47	1.92	0.03
Hard-Easy, Trial 1	18.26	1.74	1.91	0.04
Hard-Easy, Trial 2	18.50	2.43	1.96	0.04
Hard-Easy, Trial 3	18.88	2.59	1.93	0.04
Hard-Easy, Trial 4	18.28	2.40	1.91	0.04
Hard-Easy, Trial 5	17.84	2.26	1.94	0.04
Hard-Easy, Trial 6	17.71	2.49	1.91	0.04

Condition/Trial	Comms, % Correct		Comms, % Timeouts	
	Mean	SEM	Mean	SEM
Easy, Trial 1	98.44	1.56	1.56	1.56
Easy, Trial 2	98.44	1.56	1.56	1.56
Easy, Trial 3	98.44	1.56	1.56	1.56
Easy, Trial 4	95.31	2.52	1.56	1.56
Easy, Trial 5	96.88	2.13	1.56	1.56
Easy, Trial 6	95.31	3.40	3.13	2.13
Hard, Trial 1	96.34	1.46	2.19	1.12
Hard, Trial 2	91.78	2.23	7.61	2.15
Hard, Trial 3	92.17	2.33	6.64	2.17
Hard, Trial 4	93.31	1.84	6.17	1.92
Hard, Trial 5	91.76	3.75	6.03	3.29
Hard, Trial 6	88.38	3.53	8.79	3.36
Easy-Hard, Trial 1	91.72	2.41	8.28	2.41
Easy-Hard, Trial 2	91.22	3.88	7.29	3.75
Easy-Hard, Trial 3	94.48	2.44	4.59	2.48
Easy-Hard, Trial 4	89.85	3.60	9.22	3.70
Easy-Hard, Trial 5	92.49	3.23	6.55	2.71
Easy-Hard, Trial 6	92.90	2.24	4.13	2.15
Hard-Easy, Trial 1	93.53	2.05	3.41	1.58
Hard-Easy, Trial 2	94.80	2.39	4.28	2.32
Hard-Easy, Trial 3	88.40	5.09	10.23	4.82
Hard-Easy, Trial 4	91.49	2.98	5.90	2.86
Hard-Easy, Trial 5	92.50	2.97	5.97	3.06
Hard-Easy, Trial 6	91.26	3.82	7.59	3.82

Condition/Trial	Comms, Mean Correct RT	
	Mean	SEM
Easy, Trial 1	8.79	0.36
Easy, Trial 2	8.78	0.32
Easy, Trial 3	8.47	0.22
Easy, Trial 4	8.78	0.34
Easy, Trial 5	8.61	0.34
Easy, Trial 6	8.80	0.28
Hard, Trial 1	8.89	0.25
Hard, Trial 2	8.83	0.27
Hard, Trial 3	8.89	0.27
Hard, Trial 4	8.90	0.23
Hard, Trial 5	8.79	0.27
Hard, Trial 6	9.06	0.29
Easy-Hard, Trial 1	8.93	0.29
Easy-Hard, Trial 2	8.83	0.27
Easy-Hard, Trial 3	9.15	0.39
Easy-Hard, Trial 4	9.16	0.28
Easy-Hard, Trial 5	9.14	0.32
Easy-Hard, Trial 6	8.98	0.31
Hard-Easy, Trial 1	8.94	0.22
Hard-Easy, Trial 2	9.02	0.32
Hard-Easy, Trial 3	8.98	0.27
Hard-Easy, Trial 4	9.01	0.30
Hard-Easy, Trial 5	8.97	0.29
Hard-Easy, Trial 6	8.83	0.24

Note. Comms = Communications Task

F2. Means and standard errors for all performance measures as a function of minute and condition

Condition/Min	Systems, % correct		Systems, % False Alarms	
	Mean	SEM	Mean	SEM
Easy, Min 1	94.83	0.79	11.73	0.86
Easy, Min 2	93.90	1.00	6.94	1.13
Easy, Min 3	95.32	0.81	5.47	1.51
Easy, Min 4	95.41	0.94	4.36	1.42
Easy, Min 5	91.00	1.83	6.25	1.22
Easy, Min 6	93.85	1.81	8.95	2.20
Hard, Min 1	81.18	1.80	11.57	1.41
Hard, Min 2	77.28	2.69	14.59	1.68
Hard, Min 3	80.00	2.61	14.11	1.59
Hard, Min 4	83.64	2.42	11.78	1.51
Hard, Min 5	79.54	2.61	13.56	1.73
Hard, Min 6	81.79	2.36	11.89	1.37
Easy-Hard, Min 1	97.09	0.61	6.47	1.75
Easy-Hard, Min 2	94.21	1.27	7.61	2.02
Easy-Hard, Min 3	94.52	1.39	7.78	1.84
Easy-Hard, Min 4	80.15	2.49	10.11	1.34
Easy-Hard, Min 5	80.27	2.54	12.92	1.53
Easy-Hard, Min 6	78.67	2.72	12.60	1.63
Hard-Easy, Min 1	80.97	1.73	11.59	1.22
Hard-Easy, Min 2	76.19	2.86	14.14	1.95
Hard-Easy, Min 3	80.52	2.67	12.32	1.56
Hard-Easy, Min 4	93.86	1.29	8.82	2.08
Hard-Easy, Min 5	92.68	1.68	8.20	1.86
Hard-Easy, Min 6	95.87	1.12	7.10	2.16

Condition/Min	Systems, % timeouts		Systems, correct Mean RT	
	Mean	SEM	Mean	SEM
Easy, Min 1	5.17	0.79	1.56	0.05
Easy, Min 2	6.10	1.00	1.64	0.04
Easy, Min 3	4.68	0.81	1.70	0.06
Easy, Min 4	4.57	0.94	1.61	0.04
Easy, Min 5	9.00	1.83	1.68	0.04
Easy, Min 6	6.15	1.81	1.50	0.05
Hard, Min 1	18.82	1.80	1.97	0.03
Hard, Min 2	22.72	2.69	2.02	0.05
Hard, Min 3	20.00	2.61	1.99	0.04
Hard, Min 4	16.36	2.42	1.92	0.04
Hard, Min 5	20.46	2.61	2.02	0.04
Hard, Min 6	18.21	2.36	2.01	0.04
Easy-Hard, Min 1	2.91	0.61	1.50	0.05
Easy-Hard, Min 2	5.79	1.27	1.56	0.05
Easy-Hard, Min 3	5.48	1.39	1.61	0.05
Easy-Hard, Min 4	19.85	2.49	1.95	0.04
Easy-Hard, Min 5	19.73	2.54	1.99	0.03
Easy-Hard, Min 6	21.33	2.72	1.98	0.04
Hard-Easy, Min 1	19.03	1.73	2.01	0.03
Hard-Easy, Min 2	23.81	2.86	2.05	0.03
Hard-Easy, Min 3	19.48	2.67	1.98	0.04
Hard-Easy, Min 4	5.99	1.31	1.58	0.04
Hard-Easy, Min 5	7.32	1.68	1.58	0.04
Hard-Easy, Min 6	4.06	1.09	1.68	0.08

Condition/Min	Comms, % Correct		Comms, % Timeouts	
	Mean	SEM	Mean	SEM
Easy, Min 1	99.22	0.78	0.78	0.78
Easy, Min 2	94.69	3.01	1.25	1.25
Easy, Min 3	96.35	2.52	3.65	2.52
Easy, Min 4	98.44	1.56	0.00	0.00
Easy, Min 5	96.88	2.27	3.13	2.27
Easy, Min 6	100.00	0.00	0.00	0.00
Hard, Min 1	94.41	1.98	3.78	1.61
Hard, Min 2	93.42	1.95	5.28	1.58
Hard, Min 3	91.23	2.77	7.20	2.67
Hard, Min 4	94.32	2.33	4.60	2.17
Hard, Min 5	88.84	3.19	9.32	2.91
Hard, Min 6	93.32	2.32	5.82	2.36
Easy-Hard, Min 1	97.92	2.08	0.00	0.00
Easy-Hard, Min 2	99.22	0.78	0.78	0.78
Easy-Hard, Min 3	96.35	2.52	0.00	0.00
Easy-Hard, Min 4	92.99	1.85	6.34	1.82
Easy-Hard, Min 5	88.98	4.27	9.81	4.33
Easy-Hard, Min 6	90.21	3.25	8.49	3.27
Hard-Easy, Min 1	93.68	2.92	0.00	0.00
Hard-Easy, Min 2	91.33	3.16	4.91	2.92
Hard-Easy, Min 3	90.44	3.46	7.93	2.91
Hard-Easy, Min 4	93.46	3.61	5.96	2.71
Hard-Easy, Min 5	93.43	3.95	5.90	3.63
Hard-Easy, Min 6	95.49	3.28	3.93	2.81

Condition/Min	Comms, Mean Correct RT	
	Mean	SEM
Easy, Min 1	9.08	0.29
Easy, Min 2	8.75	0.40
Easy, Min 3	9.18	0.41
Easy, Min 4	8.35	0.34
Easy, Min 5	8.88	0.35
Easy, Min 6	8.00	0.50
Hard, Min 1	9.34	0.26
Hard, Min 2	9.16	0.31
Hard, Min 3	8.76	0.23
Hard, Min 4	8.49	0.23
Hard, Min 5	8.84	0.27
Hard, Min 6	8.75	0.29
Easy-Hard, Min 1	8.78	0.31
Easy-Hard, Min 2	8.77	0.24
Easy-Hard, Min 3	8.71	0.50
Easy-Hard, Min 4	9.28	0.30
Easy-Hard, Min 5	8.89	0.31
Easy-Hard, Min 6	8.98	0.32
Hard-Easy, Min 1	9.50	0.32
Hard-Easy, Min 2	9.06	0.29
Hard-Easy, Min 3	8.92	0.27
Hard-Easy, Min 4	7.84	0.33
Hard-Easy, Min 5	7.84	0.25
Hard-Easy, Min 6	6.98	0.28

Note. Comms = Communications Task

APPENDIX G:

ANOVA Tables

Table G1. Results of the 4 (condition) x 6 (trial) repeated measures ANOVAs

Performance Measure	Effect	df	MS	<i>F</i>	<i>p</i> levels	η_p^2
Systems, % Correct	Main effect of condition	2.29	4457.37	39.12	0.00	.723
	Error	34.36	113.93			
	Main effect of trial	1.60	60.31	0.29	0.70	.019
	Error	23.96	208.27			
	Condition*Trial	6.36	35.60	0.69	0.66	.044
	Error	95.43	51.23			
Systems, % Timeouts	Main effect of condition	2.29	4457.37	39.12	.000	.723
	Error	34.36	113.93			
	Main effect of trial	1.60	60.31	.29	.702	.019
	Error	23.96	208.27			
	Condition*Trial	6.36	35.60	.69	.663	.044
	Error	95.43	51.23			
Systems, % False Alarms	Main effect of condition	2.66	692.62	19.90	.000	.570
	Error	39.91	34.81			
	Main effect of trial	3.95	10.68	.83	.511	.052
	Error	59.26	12.89			
	Condition*Trial	6.20	23.92	.99	.439	.062
	Error	92.94	24.18			
Systems, Mean correct RT	Main effect of condition	2.38	3.13	85.87	.000	.851
	Error	35.71	.04			
	Main effect of trial	3.26	.03	1.08	.371	.067
	Error	48.90	.03			
	Condition*Trial	7.00	.04	1.28	.270	.078
	Error	105.02	.03			

Performance Measure	Effect	df	MS	<i>F</i>	<i>p</i> levels	η_p^2
Communications, % Correct	Main effect of condition	2.41	550.59	5.01	.008	.250
	Error	36.22	109.90			
	Main effect of trial	2.83	149.73	1.85	.155	.110
	Error	42.40	80.90			
	Condition*Trial	2.87	415.39	1.34	.274	.082
	Error	43.03	309.91			
Communications, % Timeouts	Main effect of condition	2.41	628.30	3.49	.034	.189
	Error	36.09	180.15			
	Main effect of trial	2.65	70.27	.81	.484	.051
	Error	39.73	87.05			
	Condition*Trial	3.49	301.06	1.57	.202	.095
	Error	52.30	191.81			
Communications, Mean correct RT	Main effect of condition	2.66	2.11	4.32	.012	.224
	Error	39.97	.49			
	Main effect of trial	3.48	.13	.33	.833	.021
	Error	52.26	.39			
	Condition*Trial	6.13	.60	.82	.561	.052
	Error	92.01	.73			
Resource Management, Average deviation from target	Main effect of condition	2.44	277694.87	7.84	.001	.343
	Error	36.58	35410.01			
	Main effect of trial	2.75	55390.97	4.23	.013	.220
	Error	41.27	13090.74			
	Condition*Trial	4.55	39918.75	1.48	.211	.090
	Error	68.22	26914.31			
Tracking RMS error	Main effect of condition	1.65	132764.94	148.09	0.00	.908
	Error	24.71	896.50			
	Main effect of trial	3.51	1042.50	9.30	0.00	.383
	Error	52.72	112.07			
	Condition*Trial	6.28	192.37	1.39	0.22	.085
	Error	94.23	138.04			

Note. RT = reaction time

Table G2. Results of the 4 (condition) x 6 (minute) repeated measures ANOVAs

Performance Measure	Effect	df	MS	<i>F</i>	<i>p</i> levels	η_p^2
	Main effect of condition	2.2	481.45	3.67	0.03	0.196
	Error	32.95	131.34			
	Main effect of minute	1.8	676.92	3.23	0.06	0.177
	Error	27.03	209.81			
	Condition * Minute	3.76	397.98	2.87	0.03	0.161
	Error	56.41	138.52			
Communications, % Timeouts	Main effect of condition	2.11	8.86	4.75	0.01	0.241
	Error	31.71	1.87			
	Main effect of minute	3.07	13.34	10.66	0	0.415
	Error	46	1.25			
	Condition * Minute	4.29	13.77	6.68	0	0.308
	Error	64.29	2.06			
Communications, Mean correct RT	Main effect of condition	2.69	215686.7	11.29	0	0.43
	Error	40.35	19096.67			
	Main effect of minute	3.12	26846.06	2.6	0.06	0.148
	Error	46.75	10325.05			
	Condition * Minute	4.65	93779.44	5.19	0	0.257
	Error	69.78	18063.47			
Tanks: Average deviation from target	Main effect of condition	1.55	138016.07	149.31	0	0.909
	Error	23.28	924.33			
	Main effect of minute	3.08	2382.32	18.72	0	0.555
	Error	46.13	127.28			
	Condition * Minute	2.59	71993.23	166.28	0	0.917
	Error	38.9	432.97			
Tracking RMS error						

Note. RT = reaction time

Table G3. Results of the one-way ANOVAs, examining main effect of condition, for each minute

Performance Measure	Minute	df	MS	<i>F</i>	<i>p</i> levels	η_p^2
Systems, % Correct	Minute 1	1.83	1960.88	79.68	.000	.842
	Error	27.42	24.61			
	Minute 2	1.79	2684.96	46.72	.000	.757
	Error	26.86	57.47			
	Minute 3	2.50	1380.46	26.83	.000	.641
	Error	37.43	51.45			
	Minute 4	2.32	1170.06	30.54	.000	.671
	Error	34.80	38.32			
	Minute 5	1.65	1399.60	30.67	.000	.672
	Error	24.72	45.64			
	Minute 6	2.59	1361.94	31.24	.000	.676
	Error	38.92	43.60			
Systems, % Timeouts	Minute 1	1.83	1960.88	79.68	.000	.842
	Error	27.42	24.61			
	Minute 2	1.79	2684.96	46.72	.000	.757
	Error	26.86	57.47			
	Minute 3	2.50	1380.46	26.83	.000	.641
	Error	37.43	51.45			
	Minute 4	2.35	1166.95	30.17	.000	.668
	Error	35.31	38.68			
	Minute 5	1.65	1399.60	30.67	.000	.672
	Error	24.72	45.64			
	Minute 6	2.60	1364.10	31.22	.000	.675
	Error	39.05	43.69			

Performance Measure	Minute	df	MS	<i>F</i>	<i>p</i> levels	η_p^2
Systems, % False Alarms	Minute 1	2.24	142.24	7.66	.001	.338
	Error	33.66	18.58			
	Minute 2	2.38	340.06	22.05	.000	.595
	Error	35.71	15.42			
	Minute 3	2.21	344.33	13.06	.000	.465
	Error	33.20	26.37			
	Minute 4	2.47	196.41	10.15	.000	.404
	Error	37.09	19.35			
	Minute 5	1.98	309.21	18.86	.000	.557
	Error	29.71	16.40			
	Minute 6	1.82	174.78	5.02	.016	.251
	Error	27.24	34.81			
Systems, Mean correct RT	Minute 1	2.08	1.61	67.67	.000	.819
	Error	31.27	.02			
	Minute 2	2.36	1.29	52.69	.000	.778
	Error	35.33	.02			
	Minute 3	2.78	.64	23.47	.000	.610
	Error	41.77	.03			
	Minute 4	2.82	.64	46.14	.000	.755
	Error	42.26	.01			
	Minute 5	1.81	1.32	59.14	.000	.798
	Error	27.08	.02			
	Minute 6	2.00	1.43	38.67	.000	.721
	Error	30.07	.04			

Performance Measure	Minute	df	MS	<i>F</i>	<i>p</i> levels	η_p^2
Comms, % Timeouts	Minute 1	1.33	117.22	3.88	.053	.205
	Error	19.91	30.21			
	Minute 2	1.53	175.78	3.79	.048	.202
	Error	22.97	46.33			
	Minute 3	2.71	235.22	3.56	.026	.192
	Error	40.72	66.13			
	Minute 4	1.89	215.95	2.62	.093	.149
	Error	28.30	82.35			
	Minute 5	2.18	216.63	2.25	.117	.130
	Error	32.68	96.25			
	Minute 6	2.34	260.93	3.98	.022	.210
	Error	35.15	65.61			
Comms, Mean correct RT	Minute 1	2.51	1.86	6.34	.002	.297
	Error	37.70	.29			
	Minute 2	2.05	.97	1.51	.238	.233
	Error	30.74	.65			
	Minute 3	2.00	1.09	.69	.511	.091
	Error	30.02	1.58			
	Minute 4	2.39	7.08	9.50	.000	.163
	Error	35.91	.75			
	Minute 5	2.49	5.17	7.66	.001	.044
	Error	37.29	.67			
	Minute 6	1.56	25.10	9.54	.002	.696
	Error	23.39	2.63			

Performance Measure	Minute	df	MS	<i>F</i>	<i>p</i> levels	η_p^2
Tracking RMS error	Minute 1	2.13	26888.33	229.63	.000	.939
	Error	31.99	117.09			
	Minute 2	2.13	36661.18	157.89	.000	.913
	Error	31.94	232.19			
	Minute 3	1.71	47450.43	136.02	.000	.901
	Error	25.59	348.85			
	Minute 4	2.24	23020.33	163.77	.000	.916
	Error	33.54	140.56			
	Minute 5	1.61	36366.74	138.85	.000	.903
	Error	24.21	261.91			
	Minute 6	1.67	44430.65			
	Error	25.11	287.42	154.58	.000	.912
Resource Management, average deviation from target	Minute 1	1.84	57435.67	6.07	.008	.288
	Error	27.54	9469.22			
	Minute 2	2.29	99411.68	9.61	.000	.390
	Error	34.40	10349.31			
	Minute 3	2.25	88367.98	6.78	.002	.311
	Error	33.74	13026.72			
	Minute 4	2.29	25453.73	3.62	.032	.194
	Error	34.28	7030.21			
	Minute 5	2.15	84042.31	7.87	.001	.344
	Error	32.18	10676.28			
	Minute 6	1.41	174313.85	9.45	.003	.386
	Error	21.15	18447.91			

Note. Comms = Communications task; RT = reaction time

Table G4. Result of one-way ANOVA for NASA TLX subscales

Subscale	df	MS	<i>F</i>	<i>p</i> levels	η_p^2
Mental Demand	1.41	29700.01	93.58	0.000	.862
Error	21.15	317.38			
Physical Demand	1.09	23071.18	31.95	0.000	.681
Error	16.41	721.99			
Performance	1.70	3465.83	33.25	0.000	.689
Error	25.45	104.24			
Temporal Demand	1.28	31886.08	79.13	0.000	.841
Error	19.18	402.98			
Effort	1.31	19170.12	41.14	0.000	.733
Error	19.63	466.01			
Frustration	1.52	8798.17	35.11	0.000	.701
Error	22.87	250.61			

APPENDIX H:

Result of Tukey's HSD tests for percent correct in the Systems task

Table H1. Means in the Systems task for percent correct, across minutes and conditions

	Minute 1	Minute 2	Minute 3	Minute 4	Minute 5	Minute 6
Easy	94.83	93.90	95.32	95.41	91.00	93.85
Difficult	81.18	77.28	80.00	83.64	79.54	81.79
Easy-difficult	97.09	94.21	94.52	80.15	80.27	78.67
Difficult-easy	80.97	76.19	80.52	93.86	92.68	95.87

Note. n = 16

Table H2. The Honestly Significant Differences, at level .01 and .05, needed and the mean squared error for each minute in the Systems task for percent correct

	Min 1	Min 2	Min 3	Min 4	Min 5	Min 6
MSerror =	15.00	34.31	42.80	29.63	25.07	37.71
HSD .01 =	4.51	6.82	7.62	6.34	5.83	7.15
HSD .05 =	3.66	5.54	6.18	5.14	4.73	5.80

Note. Critical q.01 = 4.66. Critical q.05 = 3.78

HSD = q * sqrt(MSerror / n)

Table H3. Differences between means, in each condition, for percent correct in the Systems task during each minute

	DE Min1	Difficult Min1	Easy Min1	ED Min1
DE min1	---	0.21	13.86**	16.11**
Difficult Min1		---	13.65**	15.91**
Easy min1			---	2.26
ED min1				---

	DE Min2	Difficult Min2	Easy Min2	ED Min2
DE Min2	---	1.09	17.71**	18.01**
Difficult Min2		---	16.62**	16.92**
Easy Min2			---	0.31
ED Min2				---

	Difficult Min3	DE Min3	ED Min3	Easy Min3
Difficult Min3	---	0.52	14.52**	15.32**
DE Min3		---	14.00**	14.80**
ED Min3			---	0.80
Easy Min3				---

	ED Min4	Difficult Min4	DE Min4	Easy Min4
ED Min4	---	3.49	13.71**	15.27**
Difficult Min4		---	10.22**	11.77**
DE Min4			---	1.55
Easy Min4				---

	Difficult Min5	ED Min5	Easy Min5	DE Min5
Difficult Min5	---	0.73	11.47**	13.14**
ED Min5		---	10.73**	12.41**
Easy Min5			---	1.68
DE Min5				---

	ED Min6	Difficult Min6	Easy Min6	DE Min6
ED Min6	---	3.12	15.17**	17.20**
Difficult Min6		---	12.06**	14.08**
Easy Min6			---	2.03
DE Min6				---

Note. ED = Easy Difficult; DE = Difficult-easy
 * = significance .05; ** = Significance .01

APPENDIX I:

Result of Tukey's HSD tests for percent timeouts in the Systems task

Table II. Means in the Systems task for percent timeouts, across minutes and conditions

	Minute 1	Minute 2	Minute 3	Minute 4	Minute 5	Minute 6
Easy	5.17	6.10	4.68	4.57	9.00	6.15
Difficult	18.82	22.72	20.00	16.36	20.46	18.21
Easy-difficult	2.91	5.79	5.48	19.85	19.73	21.33
Difficult-easy	19.03	23.81	19.48	5.99	7.32	4.06

Note. n = 16

Table I2. The Honestly Significant Differences, at level .01 and .05, needed and the mean squared error for each minute in the Systems task for percent timeouts.

	Minute 1	Minute 2	Minute 3	Minute 4	Minute 5	Minute 6
MSerror =	15.00	34.31	42.80	30.35	25.07	37.91
HSD .01 =	4.51	6.82	7.62	6.42	5.83	7.17
HSD .05 =	3.66	5.54	6.18	5.21	4.73	5.82

Note. Critical q.01 = 4.66. Critical q.05 = 3.78
 $HSD = q * \sqrt{MSerror / n}$

Table I3. Differences between means, in each condition, for percent timeouts in the Systems task during each minute

	ED Min1	Easy Min1	Difficult Min1	DE Min1
ED Min1	---	2.26	15.91**	16.11**
Easy Min1		---	13.65**	13.86**
Difficult Min1			---	0.21
DE Min1				---

	ED Min2	Easy Min2	Difficult Min2	DE Min2
ED Min2	---	0.31	16.92**	18.02**
Easy Min2		---	16.62**	17.71**
Difficult Min2			---	1.09
DE Min2				---

	Easy Min3	ED Min3	DE Min3	Difficult Min3
Easy Min3	---	0.80	14.80**	15.32**
ED Min3		---	14.00**	14.52**
DE Min3			---	0.52
Difficult Min3				---

	Easy Min4	DE Min4	Difficult Min4	ED Min4
Easy Min4	---	1.43	11.80**	15.29**
DE Min4		---	10.37**	13.86**
Difficult Min4			---	3.49
ED Min4				---

	DE Min5	Easy Min5	ED Min5	Difficult Min5
DE Min5	---	1.68	12.41**	13.14**
Easy Min5		---	10.73**	11.47**
ED Min5			---	0.73
Difficult Min5				---

	DE Min6	Easy Min6	Difficult Min6	ED Min6
DE Min6	---	2.09	14.15**	17.26**
Easy Min6		---	12.06**	15.17**
Difficult Min6			---	3.12
ED Min6				---

Note. ED = Easy Difficult; DE = Difficult-easy
 * = significance .05; ** = Significance .01

APPENDIX J:

Results of Tukey's HSD tests for mean percent false alarms in the Systems task

Table J1. Mean percent false alarms in the Systems task across minutes and conditions

	Minute 1	Minute 2	Minute 3	Minute 4	Minute 5	Minute 6
Easy	11.73	6.94	5.47	4.36	6.25	8.95
Difficult	11.57	14.59	14.11	11.78	13.56	11.89
Easy-difficult	6.47	7.61	7.78	10.11	12.92	12.60
Difficult-easy	11.59	14.14	12.32	8.82	8.20	7.10

Note. n = 16

Table J2. The Honestly Significant Differences, at level .01 and .05, needed and the mean squared error for each minute in Systems task for percent false alarms

	Minute 1	Minute 2	Minute 3	Minute 4	Minute 5	Minute 6
MSError =	13.90	12.24	19.46	15.95	10.83	21.07
HSD .01 =	4.34	4.08	5.14	4.65	3.83	5.35
HSD .05 =	3.52	3.31	4.17	3.77	3.11	4.34

Note. Critical q.01 = 4.66. Critical q.05 = 3.78

HSD = q * sqrt(MSError / n)

Table J3. Differences between means, in each condition, for percent false alarms in the Systems task during each minute

	ED Min1	Difficult Min1	DE Min1	Easy Min1
ED Min1	---	5.10**	5.11**	5.26**
Difficult Min1		---	0.02	0.16
DE Min1			---	0.14
Easy Min1				---

	Easy Min2	ED Min2	DE Min2	Difficult Min2
Easy Min2	---	4.46	7.20**	7.66**
ED Min2		---	6.53**	6.98**
DE Min2			---	1.67
Difficult Min2				---

	Easy Min3	ED Min3	DE Min3	Difficult Min3
Easy Min3	---		6.85**	8.64**
ED Min3		---	4.54*	6.32**
DE Min3			---	1.79
Difficult Min3				---

	Easy Min4	DE Min4	ED Min4	Difficult Min4
Easy Min4	---	4.46*	5.76**	7.42**
DE Min4		---	1.30	2.96
ED Min4			---	1.67
Difficult Min4				---

	Easy Min5	DE Min5	ED Min5	Difficult Min5
Easy Min5	---	1.94	6.67**	7.31**
DE Min5		---	4.72**	5.36**
ED Min5			---	0.64
Difficult Min5				---

	DE Min6	Easy Min6	Difficult Min6	ED Min6
DE Min6	---	1.85	4.80*	5.51**
Easy Min6		---	2.95	3.66
Difficult Min6			---	0.71
ED Min6				---

Note. ED = Easy Difficult; DE = Difficult-easy
 * = significance .05; ** = Significance .01

APPENDIX K:

Results of Tukey's HSD tests for mean correct reaction time in the Systems task

Table K1. Means in the Systems task for mean correct reaction time, across minutes and conditions

	Minute 1	Minute 2	Minute 3	Minute 4	Minute 5	Minute 6
Easy	1.56	1.64	1.70	1.61	1.68	1.50
Difficult	1.97	2.02	1.99	1.92	2.02	2.01
Easy-difficult	1.50	1.56	1.61	1.95	1.99	1.98
Difficult-easy	2.01	2.05	1.98	1.58	1.58	1.68

Note. n = 16

Table K2. The Honestly Significant Differences, at level .01 and .05, needed and the mean squared error for each minute in the Systems task for mean correct reaction time.

	Minute 1	Minute 2	Minute 3	Minute 4	Minute 5	Minute 6
MSerror =	.02	.02	.03	.01	.01	.02
HSD .01 =	.15	.16	.19	.13	.14	.18
HSD .05 =	.12	.13	.15	.11	.11	.15

Note. Critical q.01 = 4.66. Critical q.05 = 3.78

HSD = q * sqrt(MSerror / n)

Table K3. Differences between means, in each condition, for mean correct reaction time in the Systems task during each minute

	ED Min1	Easy Min1	Difficult Min1	DE Min1
ED Min1	---	0.06	0.46**	0.50**
Easy Min1		---	0.41**	0.45**
DifficultMin1			---	0.04
DE Min1				---

	ED Min2	Easy Min2	Difficult Min2	DE Min2
ED Min2	---	0.08	0.46**	0.48**
Easy Min2		---	0.38**	0.40**
DifficultMin2			---	0.02
DE Min2				---

	ED Min3	Easy Min3	DE Min3	Difficult Min3
ED Min3	---	0.09	0.37**	0.38**
Easy Min3		---	0.28**	0.29**
DE Min3			---	0.01
DifficultMin3				---

	DE Min4	Easy Min4	Difficult Min4	ED Min4
DE Min4	---	0.03	0.34**	0.36**
Easy Min4		---	0.31**	0.33**
DifficultMin4			---	0.03
ED Min4				---

	DE Min5	Easy Min5	ED Min5	Difficult Min5
DE Min5	---	0.10	0.41**	0.44**
Easy Min5		---	0.32**	0.35**
ED Min5			---	0.03
DifficultMin5				---

	Easy Min6	DE Min6	ED Min6	Difficult Min6
Easy Min6	---	0.17*	0.48**	0.51**
DE Min6		---	0.30**	0.33**
ED Min6			---	0.03
DifficultMin6				---

Note. ED = Easy Difficult; DE = Difficult-easy
 * = significance .05; ** = Significance .01

APPENDIX L:

Results of Tukey's HSD tests for mean RMS error in the Tracking task

Table L1. Mean Tracking RMS error, across minutes and conditions

	Minute 1	Minute 2	Minute 3	Minute 4	Minute 5	Minute 6
Easy	59.40	66.37	62.69	62.12	71.09	63.35
Difficult	117.10	135.32	136.39	127.15	135.43	133.16
Easy-difficult	60.41	68.00	63.99	119.75	130.42	132.93
Difficult-easy	122.22	138.63	132.42	72.85	73.88	66.45

Note. n = 16

Table L2. The Honestly Significant Differences, at level .01 and .05, needed and the mean squared error for each minute in the Tracking task

	Minute 1	Minute 2	Minute 3	Minute 4	Minute 5	Minute 6
MSerror =	83.25	164.79	198.34	104.77	140.93	160.37
HSN .01 =	10.63	14.96	16.41	11.92	13.83	14.75
HSN .05 =	8.62	12.13	13.31	9.67	11.22	11.97

Note. Critical q.01 = 4.66. Critical q.05 = 3.78

$$\text{HSD} = q * \sqrt{\text{MSerror} / n}$$

Table L3. Differences between means, in each condition, for mean RMS error in the Tracking task

	Easy Min1	ED Min1	Difficult Min1	DE Min1
Easy Min1	---	0.01	57.70**	62.82**
ED Min1		---	56.70**	61.81**
Difficult Min1			---	5.12
DE Min1				---

	Easy Min2	ED Min2	Difficult Min2	DE Min2
Easy Min2	---	1.63	68.96**	72.27**
ED Min2		---	67.33**	70.64**
Difficult Min2			---	3.31
DE Min2				---

	Easy Min3	ED Min3	DE Min3	Difficult Min3
Easy Min3	---	1.30	69.73**	73.70**
ED Min3		---	68.43**	72.40**
DE Min3			---	3.97
Difficult Min3				---

	Easy Min4	DE Min4	ED Min4	Difficult Min4
Easy Min4	---	10.737*	57.64**	65.034**
DE Min4		---	46.90**	54.30**
ED Min4			---	7.40
Difficult Min4				---

	Easy Min5	DE Min5	ED Min5	Difficult Min5
Easy Min5	---	2.80	59.33**	64.33**
DE Min5		---	56.54**	61.54**
ED Min5			---	5.00
Difficult Min5				---

	Easy Min6	DE Min6	ED Min6	Difficult Min6
Easy Min6	---	3.10	69.58**	69.80**
DE Min6		---	66.48**	66.71**
ED Min6			---	0.23
Difficult Min6				---

Note. ED = Easy Difficult; DE = Difficult-easy
 * = significance .05; ** = Significance .01

APPENDIX M:

**Results of the Tukey's HSD tests for mean deviation from target in the Resource
Management task**

Table M1. Mean deviations, across minutes and conditions

	Minute 1	Minute 2	Minute 3	Minute 4	Minute 5	Minute 6
Easy	389.07	371.50	375.86	379.39	375.42	360.93
Difficult	419.79	521.40	504.71	456.52	451.43	475.38
Easy-difficult	467.77	383.34	383.98	448.39	455.53	457.45
Difficult-easy	493.30	447.35	473.26	435.71	329.49	329.38

Note. n = 16

Table M2. The Honestly Significant Differences, at level .01 and .05, needed and the mean squared error for each minute

	Minute 1	Minute 2	Minute 3	Minute 4	Minute 5	Minute 6
MSerror =						
HSD .01 =	88.68	103.63	115.13	85.25	101.80	108.48
HSD .05 =	71.94	84.06	93.39	69.15	82.58	88.00

Note. Critical q.01 = 4.66. Critical q.05 = 3.78

HSD = q * sqrt(MSerror / n)

Table M3. Differences between means, in each condition, for mean deviation in the Resource Management task in each minute

	Easy Min1	ED Min1	DE Min1	Difficult Min1
Easy Min1	---	30.72	78.70*	104.24**
ED Min1		---	47.98	73.52*
DE Min1			---	25.54
Difficult Min1				---

	Easy Min2	ED Min2	DE Min2	Difficult Min2
Easy Min2	---	11.84	75.85	149.90**
ED Min2		---	64.01	138.06**
DE Min2			---	74.04
Difficult Min2				---

	Easy Min3	ED Min3	DE Min3	Difficult Min3
Easy Min3	---	8.11	97.40*	128.85**
ED Min3		---	89.28	120.73**
DE Min3			---	31.46
Difficult Min3				---

	Easy Min4	DE Min4	ED Min4	Difficult Min4
Easy Min4	---	56.32	69.00	77.13*
DE Min4		---	12.68	20.81
ED Min4			---	8.13
Difficult Min4				---

	DE Min5	Easy Min5	Difficult Min5	ED Min5
DE Min5	---	45.93	121.95**	126.04**
Easy Min5		---	76.01	80.11
Difficult Min5			---	4.10
ED Min5				---

	DE Min6	Easy Min6	ED Min6	Difficult Min6
DE Min6	---	31.54	128.07**	146.00**
Easy Min6		---	96.53*	114.45**
ED Min6			---	17.93
Difficult Min6				---

Note. ED = Easy Difficult; DE = Difficult-easy
 * = significance .05; ** = Significance .01

APPENDIX N:

Results of the Tukey's HSD tests for mean percent timeouts in the Communications task

Table N1. Mean percent timeouts in the communications task, across minutes and conditions

	Minute 1	Minute 2	Minute 3	Minute 4	Minute 5	Minute 6
Easy	.78	1.25	3.65	.00	3.13	.00
Difficult	3.78	5.28	7.20	4.60	9.32	5.82
Easy-difficult	.00	.78	.00	6.34	9.81	8.49
Difficult-easy	.00	4.91	7.93	5.96	5.90	3.93

Note. n = 16

Table N2. The Honestly Significant Differences, at level .01 and .05, needed and the Mean Squared Error for each minute in the communications task

	Minute 1	Minute 2	Minute 3	Minute 4	Minute 5	Minute 6
MSerror =			59.83			51.25
HSD .01 =			9.01			8.34
HSD .05 =			7.31			6.77

Note. Critical q.01 = 4.66. Critical q.05 = 3.78

HSD = q * sqrt(MSerror / n)

Table N3. Differences between means, in each condition, for mean percent timeouts in the Communication Task for each minute

	ED Min3	Easy Min3	Difficult Min3	DE Min3
ED Min3	---	3.65	7.20	7.93*
Easy Min3		---	3.56	4.28
Difficult Min3			---	0.72
DE Min3				---

	Easy Min6	DE Min6	Difficult Min6	ED Min6
Easy Min6	---	3.93	5.82	8.50**
DE Min6		---	1.89	4.56
Difficult Min6			---	2.67
ED Min6				---

Note. ED = Easy Difficult; DE = Difficult-easy
 * = significance .05; ** = Significance .01

APPENDIX O:

Results of the Tukey's HSD tests for mean reaction time in the Communications task

Table O1. Mean correct reaction time for the Communications task across minutes and conditions

	Minute 1	Minute 2	Minute 3	Minute 4	Minute 5	Minute 6
Easy	9.08	8.75	9.18	8.35	8.88	8.00
Difficult	9.34	9.16	8.76	8.49	8.84	8.75
Easy-difficult	8.78	8.77	8.71	9.28	8.89	8.98
Difficult-easy	9.50	9.06	8.92	7.84	7.84	6.98

Note. n = 16

Table O2. The Honestly Significant Differences, at level .01 and .05, needed and the Mean Squared Error for each minute in the Communications task

	Minute 1	Minute 2	Minute 3	Minute 4	Minute 5	Minute 6
MSerror =	.25			.59	.56	1.37
HSD .01 =	.58			.90	.87	1.36
HSD .05 =	.47			.73	.71	1.11

Note. Critical q.01 = 4.66. Critical q.05 = 3.78

HSD = q * sqrt(MSerror / n)

Table O3. Differences between means, in each condition, for mean correct reaction time in the Communication Task for each minute

	EDmin1	Easy Min1	Difficult Min1	DEmin1
EDmin1	---	0.29	0.55	0.71**
Easy Min1		---	0.26	0.42
Difficult Min1			---	0.16
DEmin1				---

	DEmin4	Easy Min4	Difficult Min4	EDmin4
DEmin4	---	0.51	0.61	1.43**
Easy Min4		---	0.14	0.93**
Difficult Min4			---	0.79*
EDmin4				---

	DEmin5	Difficult Min5	Easy Min5	EDmin5
DEmin5	---	1.00**	1.04**	1.06**
Difficult Min5		---	0.03	0.05
Easy Min5			---	0.02
EDmin5				---

	DEmin6	Easy Min6	Difficult Min6	EDmin6
DEmin6	---	1.03	1.78**	2.00**
Easy Min6		---	0.75	0.98
Difficult Min6			---	0.23
EDmin6				---

Note. ED = Easy Difficult; DE = Difficult-easy

* = significance .05; ** = Significance .01

APPENDIX P:

Results of Tukey's HSD tests for mean TLX subscale scores for each condition

Table P1. Mean TLX subscale scores for each condition

	Ment. Demand	Phys. Demand	Perf.	Temp. Demand	Effort	Frust.
Easy	9.08	8.75	9.18	8.35	8.88	8.00
Difficult	9.34	9.16	8.76	8.49	8.84	8.75
Easy-difficult	8.78	8.77	8.71	9.28	8.89	8.98
Difficult-easy	9.50	9.06	8.92	7.84	7.84	6.98

Note. n = 16;

Table P2. TDE Honestly Significant Differences, at level .01 and .05, needed and tDE Mean Squared Error for each TLX subscale

	Ment. Demand	Phys. Demand	Perf.	Temp. Demand	Effort	Frust.
MSerror =	149.16	263.30	58.95	171.75	203.33	127.36
HSD .01 =	14.23	18.90	8.94	15.27	16.61	13.15
HSD .05 =	11.54	15.33	7.26	12.38	7.90	10.66

Note. Critical q.01 = 4.66.

Critical q.05 = 3.78

HSD = q * sqrt(MSerror / n)

Table P3. Differences between means, in each condition, for mean TLX subscale scores

	MDeasy	MD DE	MD ED	MDdifficult
MDeasy	---	2.19	50.73**	53.65**
MD DE		---	48.54**	51.46**
MD ED			---	2.92
MDdifficult				---

MD= Mental
Demand

	PDeasy	PD DE	PD ED	PDdifficult
PDeasy	---	0.62	37.87**	41.98**
PD DE		---	37.24**	41.35**
PD ED			---	4.11
PDdifficult				---

PD = Physical
Demand

	PerfEasy	Perf DE	Perf ED	PerfDifficult
PerfEasy	---	2.29	16.15**	23.073**
Perf DE		---	13.85**	20.78**
Perf ED			---	6.93
PerfDifficult				---

Perf =
Performance

	TDeasy	TD DE	TD ED	TDdifficult
TDeasy	---	1.56	49.22**	53.13**
TD DE		---	47.66**	51.56**
TD ED			---	3.91
TDDifficult				---

TD = Temporal
Demand

	EffortEasy	Effort DE	Effort ED	EffortDifficult
EffortEasy	---	1.41	38.39**	42.03**
Effort DE		---	36.98**	40.63**
Effort ED			---	3.65
EffortDifficult				---

	FrustEasy	Frust DE	Frust ED	FrustDifficult
FrustEasy	---	0.10	22.92**	33.23**
Frust DE		---	22.81**	33.13**
Frust ED			---	10.31
FrustDifficult				---

Frust =
Frustration

CURRICULUM VITAE

Margaret (Funke) Bowers
Research Associate

Professional Education:

University of Dayton MA: Human Factors Psychology	2005-2007 (all but thesis)
University of Cincinnati BA: Psychology	2001-2005

Career:

Graduate Internship as a Consortium Research Fellow 711 HPW/RHCP, Collaborative Interfaces Branch	2005-2008
Research Associate 711/HPW/RHPA Vulnerability Analysis Branch	2008-2010
Research Associate 711/HPW/RHCP Collaborative Interfaces Branch	2010-Present

Publications:

McKinley, R.A., McIntire, L.K., Funke, M.A. (2011). Operator selection for unmanned aerial vehicle operators: A comparison of video game players and manned aircraft pilots. *Aviation, Space, and Environmental Medicine*, 82(6), 635-642.

Caldwell, L.J., Ruth, C., Funke, M.A., McIntire, L.K., Storm, W.F., Sundstrom, J.N. (2010). Validation of tDE aggregate wakefulness and readiness estimator (AWARE) using on-tDE-job security force personnel. Technical Report AFRL-RH-WP-TR-2010-0110.

Caldwell, L.J., Schmidt, R.M., Lopez, N., Ruth, C., Funke, M.A., Parker, J.G., Kirbas, C., Zalusky, E., Sudeepa, G., Quill, L.L., Nemeth, K.J., Taylor, B. (2010). TDE Utility of fMRI for assessing and predicting individual differences in fatigue vulnerability. Technical Report AFRL-RH-WP-TR-2010-0107.

Schmidt, R.M., Robinson, T.B., Ruth, C.R., Shiyanov, P.A., Goodyear, C.D., Walters, C.M., Funke, M.A., McIntire, L.K. (2010). Facial Expressions of Emotion and the Assessment of Performance. Human Effectiveness Directorate: Biosciences and Performance Division. Technical Report AFRL-RH-WP-TR-2010-0114.

Schmidt, R. M., Wilson, G., Davis, I., & Funke, M. (2008). Assessment and classification of cognitive decrements associated with high workload and extended work periods in a UAV setting. Human Effectiveness Directorate: Biobehavioral Performance Branch. Technical Report AFRL-RH-WP-TP-2010-0015.