NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

SYSTEMS ENGINEERING CAPSTONE REPORT

AN INVESTIGATION OF MEDEVAC AIR CREW COGNITIVE WORKLOAD: USING IMPRINT TO ASSESS THE IMPACT OF MODEL GRANULARITY AND AUTOMATION

by

Matthew Brooks, Wilson Chiu, Austin Forsythe Jr.,
and Lindsay E. Keel

June 2023

Advisor: Lawrence G. Shattuck
Co-Advisor: Matthew C. Nicholson

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC, 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE June 2023	3. REPORT TYPE AND DATES COVERED Systems Engineering Capstone Report

4. TITLE AND SUBTITLE AN INVESTIGATION OF MEDEVAC AIR CREW COGNITIVE WORKLOAD: USING IMPRINT TO ASSESS THE IMPACT OF MODEL GRANULARITY AND AUTOMATION	5. FUNDING NUMBERS
6. AUTHOR(S) Matthew Brooks, Wilson Chiu, Austin Forsythe Jr., and Lindsay E. Keel	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.	12b. DISTRIBUTION CODE A

13. ABSTRACT (maximum 200 words)

The Army Future Vertical Lift (FVL) program is developing a new generation of helicopters. The Holistic Situational Awareness–Decision Making (HSA-DM) project supports the FVL program by investigating technologies that can aid in designing, developing, and integrating systems aimed at managing the cognitive workload of pilots. This study aims to determine how a model's granularity and the automation of certain tasks affects the pilot and co-pilot's cognitive workloads. Cognitive walkthroughs and interviews were conducted with five military helicopter pilots to elicit task workload values. An existing Improved Performance Research Integration Tool (IMPRINT) model from a previous study was modified by increasing its granularity, then both models were altered by automating selective tasks, resulting in four distinct models. Each model was run ten times. Average cognitive workload and workload spikes were analyzed for both the pilot and co-pilot. The average cognitive workload was highest for the pilot in the more granular model as greater model granularity contributed to higher spikes. Automation added to the more granular model resulted in the greatest reduction of pilot workload. This analysis establishes a repeatable framework for future studies.

14. SUBJECT TERMS cognitive load, cognitive overload, automation, Army Future Vertical Lift, FVL, Holistic Situational Awareness–Decision Making, HSA-DM, Improved Performance Research Integration Tool, IMPRINT	15. NUMBER OF PAGES 135
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

AN INVESTIGATION OF MEDEVAC AIR CREW COGNITIVE WORKLOAD: USING IMPRINT TO ASSESS THE IMPACT OF MODEL GRANULARITY AND AUTOMATION

MAJ Matthew Brooks (USA), MAJ Wilson Chiu (USA), MAJ Austin Forsythe Jr. (USA), and CPT Lindsay E. Keel (USA)

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN SYSTEMS ENGINEERING MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
June 2023

Lead Editor: Matthew Brooks

Reviewed by:
Lawrence G. Shattuck
Advisor

Matthew C. Nicholson
Co-Advisor

Accepted by:
Oleg A. Yakimenko
Chair, Department of Systems Engineering

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The Army Future Vertical Lift (FVL) program is developing a new generation of helicopters. The Holistic Situational Awareness–Decision Making (HSA-DM) project supports the FVL program by investigating technologies that can aid in designing, developing, and integrating systems aimed at managing the cognitive workload of pilots. This study aims to determine how a model's granularity and the automation of certain tasks affects the pilot and co-pilot's cognitive workloads. Cognitive walkthroughs and interviews were conducted with five military helicopter pilots to elicit task workload values. An existing Improved Performance Research Integration Tool (IMPRINT) model from a previous study was modified by increasing its granularity, then both models were altered by automating selective tasks, resulting in four distinct models. Each model was run ten times. Average cognitive workload and workload spikes were analyzed for both the pilot and co-pilot. The average cognitive workload was highest for the pilot in the more granular model as greater model granularity contributed to higher spikes. Automation added to the more granular model resulted in the greatest reduction of pilot workload. This analysis establishes a repeatable framework for future studies.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I. INTRODUCTION 1
- A. BACKGROUND 1
- B. STAKEHOLDERS 2
- C. PROJECT OVERVIEW 3
- D. RESEARCH QUESTIONS 4

II. COGNITIVE WORKLOAD 7
- A. MODELING AND MEASURING COGNITIVE WORKLOAD 12
- B. AUTOMATION IN FUTURE VERTICAL LIFT 17
- C. IMPROVED PERFORMANCE RESEARCH INTEGRATION TOOL 23
- D. CONCLUSION 31

III. METHODOLOGY 33
- A. OVERVIEW 33
- B. PARTICIPANTS 34
- C. EQUIPMENT 35
- D. PROCEDURES 36
  - 1. IMPRINT Model Validation and Construction 36
  - 2. Task Analysis 42
  - 3. Knowledge Elicitation 44
  - 4. Data Analysis from Interviews 46
  - 5. Updating Scheff's IMPRINT Model 46
  - 6. Automating Tasks 48

IV. RESULTS 51
- A. OVERVIEW 51
- B. COGNITIVE WORKLOAD: LOW GRANULARITY VERSUS HIGH GRANULARITY (SUBTASKS) COGNITIVE MODELS 52
- C. COGNITIVE WORKLOAD MODEL WITH AUTOMATION 55
- D. DIFFERENCE IN COGNITIVE WORKLOAD AS IT RELATES TO AUTOMATION AND GRANULARITY 57
- E. COGNITIVE WORKLOAD REDUCTION WITH AUTOMATION TO THE SYSTEM 59
- F. SUMMARY 70

V. DISCUSSION 73
A. OVERVIEW 73
B. RESEARCH QUESTION ONE: DOES COGNITIVE WORKLOAD DIFFER IF MORE GRANULARITY (SUB-TASKS) IS INCORPORATED INTO A COGNITIVE MODEL? 73
C. RESEARCH QUESTION TWO: WHAT IS THE IMPACT ON COGNITIVE WORKLOAD IF AUTOMATION IS INTRODUCED INTO A COGNITIVE MODEL? 77
D. RESEARCH QUESTION THREE: IS THERE A DIFFERENCE IN COGNITIVE WORKLOAD AS A FUNCTION OF AUTOMATION AND GRANULARITY LEVELS? 81
E. RESEARCH QUESTION FOUR: WHERE IN A MEDEVAC MISSION, AND TO WHAT EXTENT, CAN PILOT AND CO-PILOT COGNITIVE WORKLOAD BE REDUCED BY INTRODUCING AUTOMATION TO THE SYSTEM AT A SUBTASK LEVEL? 83
F. LIMITATIONS 86
G. SUMMARY 88

VI. CONCLUSION AND RECOMMENDATIONS 91
A. OVERVIEW 91
B. CONCLUSIONS 92
C. RECOMMENDATIONS FOR HSA-DM 95
D. RECOMMENDATIONS FOR FUTURE RESEARCH 96

APPENDIX. SCRIPT FOR MEDEVAC OPERATIONAL SCENARIO 99

SUPPLEMENTAL 1. CONSOLIDATED DATA OF VACP SCORES AND TASKS RECOMMENDED FOR AUTOMATION FROM SUBJECT MATTER EXPERTS 105

SUPPLEMENTAL 2. SEEING SOUNDS IMPRINT MODELS 107

LIST OF REFERENCES 109

INITIAL DISTRIBUTION LIST 113

LIST OF FIGURES

Figure 1. Performance Predictor Based on Task Demand. Source: Chen et al. (2016, 39). 8

Figure 2. The Visual Auditory Cognitive Psychomotor Workload Scale. Source: Rusnock and Borghetti (2016). 26

Figure 3. Scheff 1.0 Model Perform Hovering Flight Task 38

Figure 4. Scheff 2.0 Model Perform Hovering Flight Task with Automation 39

Figure 5. Seeing Sounds 1.0 Perform Hovering Flight Task with New Terminating Tasks 39

Figure 6. Seeing Sounds 2.0 Perform Hovering Flight Task with Automated Tasks 40

Figure 7. Land Known Landing Area Task Concept Map 41

Figure 8. Task Analysis 41

Figure 9. Level-1 IMPRINT Task Land Known Landing Area 43

Figure 10. Level-2 Land Known Landing Area Scheff 1.0 Terminating Tasks 43

Figure 11. Level-2 Land Known Landing Area Seeing Sounds 1.0 Terminating Tasks. 44

Figure 12. The Visual Auditory Cognitive Psychomotor Workload Scale. Source: Rusnock and Borghetti (2016). 45

Figure 13. Check AC Systems Scheff 1.0 Terminating Tasks 47

Figure 14. Check AC Systems Seeing Sounds 1.0 Terminating Tasks 47

Figure 15. Seeing Sounds 1.0 Operator Workload Graph 48

Figure 16. Scheff 2.0 with Automated Tasks Path Activated 49

Figure 17. Seeing Sounds 2.0 with Automated Tasks in Black 50

Figure 18. Graphic Comparison of Average VACP Scores of SCH 1.0 and SST 1.0, No Automation (10 Runs) 54

Figure 19. Graphic Comparison of VACP Scores of SCH 2.0 and SST 2.0, with Automation (10 Runs) 57

Figure 20. Graphic Comparison of Median Workload Spike Occurrences in the SCH and SST Cognitive Workload Model, Baseline versus Automated Model 58

Figure 21. Lowest SCH Workload Graph 1.0, No Automation (Run #3 of 10) 60

Figure 22. Highest SCH Workload Graph 1.0, No Automation (Run #4 of 10) 60

Figure 23. Lowest SST Workload Graph 1.0, No Automation (Run #3 of 10) 61

Figure 24. Highest SST Workload Graph 1.0, No Automation (Run #6 of 10) 62

Figure 25. Lowest SCH Workload Graph 2.0, with Automation (Run #9 of 10) 63

Figure 26. Highest SCH Workload Graph 2.0, with Automation (Run #10 of 10) 63

Figure 27. Lowest SST Workload Graph 2.0, with Automation (Run #3 of 10) 64

Figure 28. Highest SST Workload Graph 2.0, with Automation (Run #9 of 10) 65

Figure 29. Pilot Task Over 60 SST 1.0 Model (Highest Run) 66

Figure 30. Co-pilot Task Over 60 SST 1.0 Model (Highest Run) 66

Figure 31. Co-pilot Task Over 60 SCH 1.0 Model (Highest Run) 67

Figure 32. Co-pilot Task Over 60 SCH 2.0 Model (Highest Run) 68

Figure 33. Pilot Task Over 60 SST 2.0 Model (Highest Run) 69

Figure 34. Co-pilot Task Over 60 SST 2.0 Model (Highest Run) 69

Figure 35. Graphic Comparison of Average Workload Scores of SCH 1.0 and SST 1.0, No Automation (10 Runs) 74

Figure 36. Scheff 1.0 Model Perform Hovering Flight Task 75

Figure 37. Seeing Sounds 1.0 Perform Hovering Flight Task with new Terminating Tasks 76

Figure 38. Graphic Comparison of Average VACP Scores of SCH 1.0 and SST 1.0, No Automation (10 Runs) 77

LIST OF TABLES

Table 1. Overview of the Evolution of the Various IMPRINT Models 34

Table 2. Pilot Demographics..... 35

Table 3. Overview of Task Levels of the Scheff 1.0 Model..... 37

Table 4. Comparison of SCH Pilot and SCH Co-Pilot Workload, No Automation (10 Runs)..... 53

Table 5. Comparison of SST Pilot and SST Co-Pilot Workload, No Automation (10 Runs)..... 53

Table 6. Comparison of VACP Scores of SCH 1.0 and SST 1.0, No Automation (10 Runs)..... 53

Table 7. Comparison of Average Workload in SCH Cognitive Workload Model, Baseline versus Automated Mode 54

Table 8. Comparison of SCH Pilot and SCH Co-Pilot Workload, with Automation (10 Runs)..... 55

Table 9. Comparison of SST Pilot and SST Co-Pilot Workload, with Automation (10 Runs)..... 56

Table 10. Comparison of VACP Scores of SCH 2.0 and SST 2.0, with Automation (10 Runs)..... 56

Table 11. Comparison of Total and Median Workload Spike Occurrences in SCH Cognitive Workload Model, Baseline versus Automated Model 58

Table 12. Comparison of Total and Median Workload Spike Occurrences in SST Cognitive Workload Model, Baseline versus Automated Model 58

Table 13. Average Cognitive Workload, without and with automation 78

Table 14. Average Number of Workload Spikes SCH 2.0 (10 runs)..... 79

Table 15. Average Number of Workload Spikes SST 2.0 (10 runs) 80

Table 16. Comparison of Average Workload in SCH Cognitive Workload Model, Baseline versus Automated Mode 94

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AC	Aircraft
AI	Artificial Intelligence
AFC	Army Futures Command
ATM	Aircrew Training Manual
CAWP	Continuous Analytic Workload Profile
CRS	Congressional Research Service
DEVCOM	U.S. Army Combat Capabilities Development Command
EDF	European Defense Fund
EEG	Electroencephalography
FVL	Future Vertical Lift
HITL	Human-in-the-Loop
JVMF	Joint Variable Message Format
HSA-DM	Holistic Situation Awareness-Decision Making
IMPRINT	Improved Performance Research Integration Tool
MEDEVAC	Medical Evacuation
MOC	Mission Operations Center
MRT	Multiple Resource Theory
NASA-TLX	National Aeronautics and Space Task Load Index
NATO	North Atlantic Treaty Organization
NPS	Naval Postgraduate School
OOTL	Out-of-the-Loop
RPA	Remotely Piloted Aircraft
SCH	Scott Scheff IMPRINT Model
SME	Subject Matter Expert
SST	Seeing Sounds Team IMPRINT Model
STRT	Secondary Task Reaction Times
UH-60	Medium Lift Utility Helicopter – 60 Black Hawk (M)
VACP	Visual, Auditory, Cognitive, and Psychomotor

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

BACKGROUND

The Army Future Vertical Lift (FVL) program is developing a new generation of helicopters. The Congressional Research Service (CRS) describes the FVL program as "a research and development effort dedicated to discovering, investigating, and refining the technologies to provide the next generation of vertical lift aircraft for the United States Armed Forces" (Gertler, 2021, 1). The Holistic Situational Awareness – Decision Making (HSA-DM) project supports the FVL program by investigating technologies that can aid in designing, developing, and integrating systems aimed at managing the cognitive workload of pilots.

OBJECTIVE

This research contributes to the HSA-DM effort by providing comparative data between low-granularity and high-granularity cognitive workload models to determine any differences or similarities between those models concerning cognitive workload spikes and the effects of automation. The comparative analysis provided valuable data to determine if granularity affects cognitive workload during a medical evacuation (MEDEVAC) scenario.

METHODS

To accomplish this goal, we began with a literature review of cognitive workload and the methods used to measure and model this workload. We also researched how automation has been used to mitigate workload and how pilots aviate, navigate, and communicate in flight. We conducted a task analysis by leveraging the UH-60 Aircrew Training Manual (ATM), cognitive walkthroughs, and pilot interviews to identify the necessary tasks required to operate a UH-60 in a MEDEVAC scenario. We used the Improved Performance Research Integration Tool (IMPRINT) to analyze the differences and similarities between the high-granularity and low-granularity models to determine how they affected the cognitive workload spikes and if the addition of automation affected the models differently.

FINDINGS AND CONCLUSION

The team used various metrics to measure the impact of automation, including workload scores, workload spikes over 60, and VACP results. We found differences in workload values and spikes between high-granularity and low-granularity models. The workload differences between the pilots were more significant than the co-pilots. The co-pilot tasks in high granularity were added more sequentially, resulting in virtually no workload difference between the two models compared to the difference between the pilots. The overall average workload score was reduced by 8% after automation was introduced. However, the pilot's workload in the high-granularity model was significantly higher than the low-granularity model. Next, we analyzed the cognitive workload of pilots and co-pilots using two cognitive workload models, high- and low-granularity, and found that the baseline workload spike occurrences for co-pilots were higher than for pilots in the SCH model. In contrast, the opposite was true in the SST model. Automation in the SCH model increased the pilot's workload while decreasing the co-pilot's workload, but in the SST model, automation decreased the pilot's workload while increasing the co-pilot's workload. The study also found that as the granularity level increased, pilot workload spikes increased while co-pilot workload spikes decreased.

RECOMMENDATIONS FOR FUTURE RESEARCH

First, we recommend expanding the knowledge elicitation sample size and demographics (to include warrant officers and co-pilots) to obtain better consensus for task frequency, duration, and workload values. Second, we recommend removing the attention model and communication tasks from the background of the IMPRINT model and adding them as discrete tasks where necessary to allow for workload comparison between models based on multiple resource theory and VACP values. Third, we recommend designing a digital simulator that incorporates the different phases of flight and the tasks from the IMPRINT models to evaluate the impact of automation on cognitive workload. Finally, we recommend conducting additional analysis of the model where workload drops to zero and incorporating tasks identified during those periods into the model for a more realistic simulation.

References

Gertler, Jeremiah. 2021. *Army Future Vertical Lift (FVL) Program*. CRS Report No. IF11367. Washington, DC: Congressional Research Service. https://crsreports.congress.gov/product/pdf/IF/IF11367/6.

Shivers, Matthew. 2021. "Holistic Situational Awareness and Decision Making (HSA-DM)." Presentation for the Naval Postgraduate School Human Systems Integration Department, Monterey, CA, November 2021.

United States Army Aviation Center of Excellence. 2022. *UH-60 Series Aircrew Training Manual*. Fort Rucker, AL: Directorate of Training and Doctrine, Flight Training Branch.

United States Army Acquisition Support Center. "Future Vertical Lift (FVL) Family of Systems." https://asc.army.mil/web/portfolio-item/aviation_gray-eagle-uas-2/.

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

We want to express our deepest gratitude to Dr. Lawrence Shattuck and Mr. Matthew (Cody) Nicholson for your guidance, academic recommendations, critiques, and knowledge throughout our research effort. Your dedication to our team and research was integral to the success of our effort. We thank you for the mentorship and vision provided during friction points in our research. Having you both as our academic advisor team has been a privilege.

We would also like to thank Mr. Matt Shivers for allowing us the opportunity to expand upon previous cognitive workload research in support of the Army modernization effort. Additionally, we would want to recognize the commitment and significant contributions of Mr. Robert Sargent and Mr. Scott Scheff. The research effort would not have been possible without their training, insight, and willingness to guide us in understanding and building our IMPRINT model.

We also want to recognize the MEDEVAC community members who participated in our knowledge elicitation; thank you for your precious time and attention. We especially wanted to acknowledge the immense contribution of MAJ Jeffery Crook. MAJ Crook's technical expertise, input, and overall effort were pivotal to the research endeavor.

Lastly, the team wants to recognize our families' support and sacrifices. Thank you for your encouragement and support throughout our time at the Naval Postgraduate School.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. BACKGROUND

The Army Future Vertical Lift (FVL) program is developing a new generation of helicopters. The Congressional Research Service (CRS) describes the FVL program as "a research and development effort dedicated to discovering, investigating, and refining the technologies that are to provide the next generation of vertical lift aircraft for the United States Armed Forces" (Gertler 2021, 1). Gertler (2021) states that the program aims to develop technologies that improve "maneuverability, range, speed, payload, survivability, reliability, and reduced logistical footprint compared with current rotorcraft" (1). The Army intends to make a generational leap forward in aviation technology and capability by developing the new platforms and operational concepts necessary to succeed in an increasingly contested and near-peer battlespace. Core to the FVL program is the integration of state-of-the-art and disruptive innovative technologies to drive changes in the Army's helicopters, methods, and aircrew domains. The European Defense Fund (EDF) Regulation defines a disruptive technology for defense as an "enhanced or completely new technology that brings about a radical change, including a paradigm shift in the concept and conduct of defense affairs such as by replacing existing defense technologies or rendering them obsolete" (European Defense Fund [EDF] 2022, 1). Furthermore, the NATO Science and Technology Organization defines disruptive technologies as "technologies or scientific discoveries that are expected to have a major, or perhaps revolutionary, effect on NATO defense, security or enterprise functions" (North Atlantic Treaty Organization [NATO] 2022, 1). As the systems within the FVL program become increasingly complex, appropriate application of automation is paramount. There is an increasing potential for pilots to reach cognitive overload due to increasing responsibilities to manage the infusion of innovative technologies, data streams, and situation awareness tools projected to be utilized on the FVL platforms. Therefore, the FVL program must assess what tasks should be automated—and how they should be automated—to avoid pilot cognitive overload.

The Holistic Situational Awareness-Decision Making (HSA-DM) project team is part of the U.S. Army Combat Capabilities Development Command (DEVCOM) Aviation and Missile Center. It is a subordinate element of the Army Futures Command (AFC). One mission of the AFC is to research and develop new weapons systems for aircraft, missiles, and unmanned vehicles, providing one-stop technical support capability for these systems throughout their life cycles. Shivers and Morony (2021) list two goals for the FVL program:

- Provide optimized task loading for Future Vertical Lift /aviation warfighters (pilots, co- pilots, and aircrew) by developing cognitive workload management capabilities.

- Improve combat mission performance of Novice, Busy, Fatigued, and Injured pilots by delivering decision-aiding algorithms, improving human-machine interface hardware/software, and implementing autonomous flight controls. (Shivers and Morony 2021, 13)

Understanding the goals of optimization and improving combat performance for pilots has shaped our team's research questions and the general direction we have taken with our Capstone project. The research team used the Improved Performance Research Integration Tool (IMPRINT) to conduct our analysis. We then conducted a task analysis by leveraging the UH-60 Aircrew Training Manual (ATM), cognitive walkthroughs, and knowledge elicitation with expert pilots to identify the critical tasks required to operate a UH-60 in a MEDEVAC scenario. The team then expanded an existing IMPRINT model using the data gathered from our knowledge elicitation to replicate all phases of flight.

B. STAKEHOLDERS

This research addresses two primary stakeholders: Army MEDEVAC pilots and the HSA-DM project team. MEDEVAC pilots are the primary project stakeholders because they are the end users for ensuring the safe and successful operation of the FVL platforms during MEDEVAC missions. The research aims to create a high-granularity model and compare the original low-granularity model and examine how cognitive

workload is captured in both models. The goal is to determine the similarities and differences between the models and provide the stakeholders with the data and analysis to decide which model would be better suited for their needs.

Our second purpose is to advise the HSA-DM project team on task automation based on the IMPRINT model design. The HSA-DM team has identified seven operational context vignettes for their project team to consider (Shivers n.d.). This research project uses the MEDEVAC vignette for model verification, emphasizing tasks during a MEDEVAC mission.

This project focuses on the cognitive workload values of pilots and co-pilots and is exclusive to the MEDEVAC platform and its associated vignette. The cognitive workload analysis includes the execution of a MEDEVAC mission only. It does not contain an aggregate workload value over the course of an entire aircrew's deployment. Only the automation of high cognitive workload tasks was explored due to time and resource constraints. This project's results are only applicable to a MEDEVAC mission and are not expected to be generalized to the entire population of U.S. Army aviators. The results inform the sponsor of the pilots and co-pilots specifically. After comparing the models, the overall differences and similarities between the models and the cognitive workload changes, if any, are revealed when automation is added at a more granular level. The analysis can be helpful for future-generation aircraft and provide information to further refine the existing IMPRINT model. The research team agreed to only focus on one specific path within the existing model rather than attempt to provide additional detail to each path.

C. PROJECT OVERVIEW

The team used a mixed-methods research approach for this project. We first conducted an operations task analysis and identified the tasks performed by UH-60 aircrews during a MEDEVAC mission. Our initial task list was based on an existing IMPRINT model provided to our team. The team utilized the tasks in the model as the starting point for knowledge elicitation. Next, the team conducted further task analysis with UH-60 pilots to create an additional level of detail (subtasks) to the low-granularity

model and expanded the sample size used to compare the models. After receiving these inputs, our team worked with IMPRINT subject matter experts to update the low-granularity model with the new tasks and workload scores.

IMPRINT is a trade-off analysis tool that can help evaluate cognitive workload and test alternate system-crew function and task allocation, such as the introduction of automation (Sargent 2022). First, the team simulated the MEDEVAC flight and developed a baseline cognitive workload score for the pilot and co-pilot with the updated model. Then, after examining where the workload spikes were the highest, we identified which tasks were being executed by the crew during those spikes and compared the data with the low-granularity model. To conduct a "apples to apples" comparison, we selected a small number of subtasks that could be automated based on what was already automated in the low-granularity model. We then updated these subtasks in the automated model, which allowed the research team to directly compare the new workload values of the aircrew with the introduction of automated tasks against the existing model and the higher granularity model. The team then performed statistical analysis to compare the low-granularity model with the high-granularity model and the low- granularity model with automation with the high-granularity model with automation.

D. RESEARCH QUESTIONS

The overarching research questions are: Does cognitive workload differ if more granularity (subtasks) is incorporated into a cognitive model? What is the impact on cognitive workload if automation is introduced into a cognitive model? Is there a difference in cognitive workload as a function of automation and granularity levels? Where in a MEDEVAC mission, and to what extent, can pilot and co-pilot cognitive workload be reduced by introducing automation to the system at a subtask level?

Our capstone project has the following deliverables:

1. An updated IMPRINT model based on the model provided to us and an updated cognitive workload analysis of the updated model. The updated IMPRINT model has automated tasks deemed the most beneficial and feasible in reducing cognitive workload. The updated cognitive workload

analysis compares workload values between a non-automated flight and a flight that incorporates some automated tasks. In addition, the analysis shows if the introduction of automated tasks has a meaningful impact on the pilot and co-pilot cognitive workload during a MEDEVAC flight.

2. An updated list of terminating tasks was added to the Scott Scheff model to provide the additional layer of detail the team added to fully evaluate the workload demand on the crew.

3. The tasks recommended for automation are based on IMPRINT workload score impacts.

4. The team's recommendation for the level of detail necessary in future IMPRINT models to accurately depict pilots' cognitive workload during flight.

5. The team's final report will include collecting all of the data and IMPRINT models for the stakeholder for additional review and possible future research efforts. The team will conduct an out brief with the final report to the stakeholders and sponsor.

6. All the data collected.

Overall, this study will inform the FVL community, specifically the HSA-DM team, as they continue to develop their understanding of where automation can assist pilots. To provide actionable recommendations, our team will present critical works we've complied to aid in our study. The team's deliverables will be shaped by the research, techniques, and findings described in the literature review.

THIS PAGE INTENTIONALLY LEFT BLANK

II. COGNITIVE WORKLOAD

In this section, the research team explores the origin and expansion of the theory of cognitive workload from the initial studies on mental capacity to the present-day understanding of the field. Additionally, the research team will compare a compiled list of the various definitions of cognitive workload and cognitive overload found during the literature review to enhance our understanding of the two terms and their relationship. A review of the literature regarding cognitive workload is essential to the team for understanding the impact of cognitive workload on human performance while conducting multiple tasks simultaneously. Our research will provide the team with the insight to recognize the effects of workload on pilots when completing the tasks associated with flying. The preliminary theories of cognitive workload's effects on human performance and attention were developed in the late 1950s and 1960s. In 1958, Broadbent conceptualized some of the foundational theoretical concepts of human attention and mental limitations to cognitively processing data in his book Perception and Communication. His theory predicted that people could only process input from, and allocate resources to, one source at a time, and that all other messages or signals were placed into a short-term memory store until they were processed or discarded. Broadbent's theory became known as the "bottlenecking" or "filter" theory and served as one of the first models of cognitive workload or multitask interference until the 1970s (Eysenck 1994).

During the 1970s, various models were developed, advancing the concept of resource sharing, which countered Broadbent's works and suggested that humans retained a pool of attentive resources that could be allocated across different tasks as required (Wickens 2002). Various experiments were conducted supporting these studies that measured the negative effects of secondary task performance when cognitive demand for the primary task was high, proving that humans could multitask with potential limitations. Kahneman (1973) introduced that people had finite mental resources to complete tasks. He posited that the more demanding tasks consume proportional amounts of a person's cognitive "capacity" or "resources." Moray's findings (1979) advanced

Kahneman's theory and determined a close relationship between mental workload and the mental pool of resources available for people to perform multiple complex tasks. He found that cognitive overload would occur in operators when more demanding tasks exceeded a person's mental resources available to complete them.

Moray was further able to link the depletion of a person's primary resources to a reduction in the performance of secondary tasks. Figure 1 models Moray's findings and represents the relationship between mental capacity, high-level task demand, and the prediction of performance when a person's mental resources have been expended and they reach the point of cognitive overload. The figure effectively depicts the significant decrease in the mental performance of the operator (dotted line) after the workload demand exceeds the mental resource supply curve (dashed line). The research team needs to understand Morays' theory because one of our research goals will be to identify where workload exceeds mental capacity and implement automation to prevent a significant decrease in operator performance.

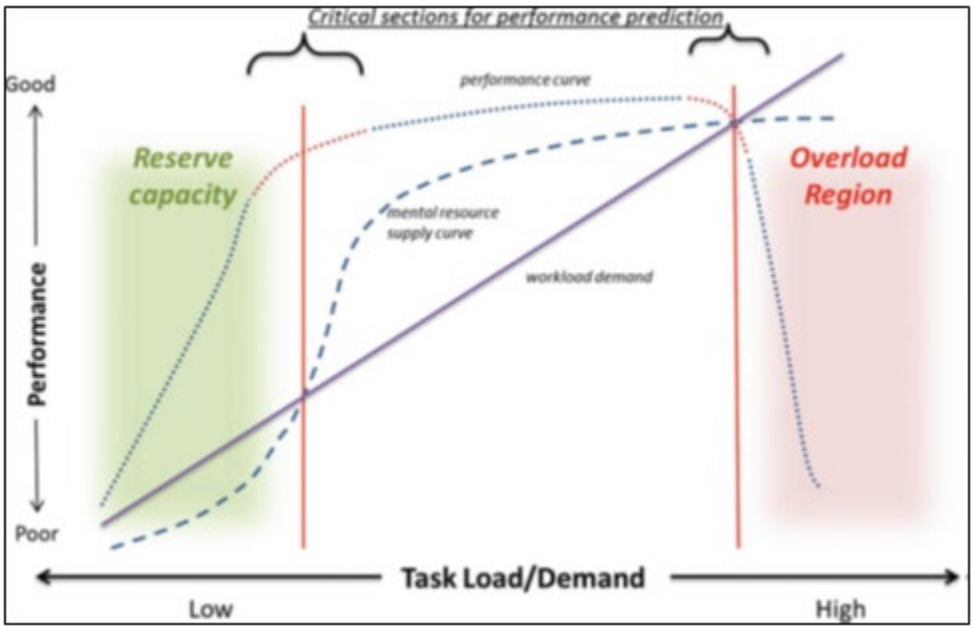

Figure 1. Performance Predictor Based on Task Demand.
Source: Chen et al. (2016, 39).

The early findings by these researchers and others established the conceptual groundwork for understanding the human capacity to process information and the effects of high-demand tasks on human performance. Understanding the impact of cognitive workload on pilots is more critical now than ever with the anticipated delivery of highly sophisticated aviation platforms and the advent of human-machine systems designed to maximize performance. In the new operational environment of machine-augmented systems, integrating human and machine systems will be critical to maximizing the effectiveness of FVL platforms. Adhikari (2015) highlighted that a person's inability to maintain high mental acuity and focus is often the point of failure when integrating humans and machines. Determining where humans reach their cognitive limitations will be instrumental in identifying the proper time to introduce machine augmentations to mitigate the effects of multitasking on pilots. As the next generation of complex human-machine systems is introduced into the future operational environment, systems operators, especially pilots, will be required to perform a greater number of tasks when compared to the current fleet of aircraft. The additional increase in instrumentation and capabilities will likely exacerbate the cognitive workload that pilots are currently experiencing, which is already triggering cognitive overload (Galant, Zawada, and Maciejewska 2020).

Our review of the theory revealed that the terms cognitive workload and cognitive overload were utilized throughout the various articles and appeared to be used interchangeably. In their research, Chen et al. (2016) highlighted that although the researchers generally agree on the fundamentals of cognitive workload, they have different interpretations of the precise definition of cognitive workload and terms associated with the phenomena. Cognitive workload, mental effort, effort, cognitive effort, mental load, and cognitive overload have been used interchangeably across various research domains (Chen et al. 2016). To better understand the concepts of cognitive workload and cognitive overload, an essential aspect of our literature review was determining how the field researchers defined the terms.

Our research reviewed several definitions of cognitive workload and cognitive overload in order to distinguish them from one another. The following definitions of cognitive workload were consolidated from relevant articles.

- Moray (Moray 1979) defined cognitive workload as "the general pool of processing resources available to all tasks and allocated to each according to its difficulty and demanded level of performance" (79).

- Johannsen (Johannsen 1979) defined mental workload as "the simultaneous effects of the input load, operator effort, and performance" (3).

- In "Multiple Resource and Performance Prediction" Wickens (2002) offered the following definition of cognitive workload, "that mental resources from a limited source and can be allocated as necessary to meet the task demands defined jointly by the level of difficulty of the task and the level of performance required" (161). He additionally included that "if a task demands more resources, it will interfere more with a concurrent task" (161).

- Chen et al. (2016) defined cognitive workload as "the load experienced by working memory when humans engage in a variety of cognitively intensive tasks ranging from learning to decision-making and more" (34).

Cognitive workload, mental effort, effort, cognitive effort, and mental load are all terms that represent the cognitive stress applied to an operator when being asked to multitask. Although various definitions of cognitive workload were found in the literature, each description considered the effect of external tasks on people when multitasking and the finite pool of available resources to process those tasks.

The following definitions do not all specifically name "cognitive overload," but all focus on the point when one's cognitive ability is overwhelmed by the number of concurrent high-demand tasks.

- Adhikari (2015) defined cognitive overload as "the state when the work task is beyond the cognitive ability of the worker" (132).
- Fox, Park, and Lang (2007) offered that "the resources required to encode the message exceed the resources allocated, and thus there are insufficient resources available for encoding, and recognition suffers" (281).
- Wickens and Alexander (2009) defined cognitive overload as "the allocation of attention to a particular channel of information, diagnostic hypothesis, or task goal, for a duration that is longer than optimal, given the expected cost of neglecting to perform other tasks" (182).
- Galant, Zawada, and Maciejewska (2020) stated that "when the maximum consumption of resources is achieved, the cognitive system that deals with coding, processing, and task allocation cannot compensate for tasks with available resources" (330).

Given these definitions, we define cognitive overload as a state of degraded mental performance due to workload demand exceeding the mental resources available. Cognitive overload occurs when the cognitive workload exceeds the operator's mental capacity to apply additional cognitive resources to complete the task (Moray 1979; Wickens 2002; Fox, Park, and Lang 2007; Wickens and Alexander 2009; Adhikari 2015; Chen et al. 2016; Galant, Zawada, and Maciejewska 2020).

The concept of task demand and the limited availability of the pool of human resources directly relates to the research endeavor. Regarding pilots, the negative results of a high-demand task environment on the crew are a depletion of the mental resources available to perform the primary functions of aviate, navigate, and communicate the helicopter (Federal Aviation Administration 2022). The depletion of attentional resources increases pilot error, and pilots' inability to maintain a high level of mental acuity for sustained periods results in aviation accidents (Galant, Zawada, and Maciejewska 2020). One of our goals as a research group is to determine where machine augmentations can

be introduced to either curtail the adverse effects of cognitive overload or enhance the pool of available resources by machines completing tasks for humans.

A. MODELING AND MEASURING COGNITIVE WORKLOAD

Our team explored various methods of collecting, analyzing, and modeling data to determine the best practices and models suited for our research. This section briefly describes the techniques and models implemented in previous research to collect and analyze data to discern the cognitive workload of various tasks or missions. The models we considered ranged from a simple statistical analysis approach based on data collection to a resource-intense process requiring specialized training, specialized equipment, and access to the monetary resources needed to conduct the research. The previous research group of Carter et al. (2021) recommended a varied approach that blended both qualitative and quantitative techniques for measuring the effects of cognitive workload; the sections below are examples of these methods that our team researched.

Data collection by surveys, instrumentation, or observation is the foundation of conducting research. During our literature review regarding cognitive workload, data collection of some magnitude was performed in all reviewed articles employing either physiological or subjective data collection techniques. Research groups used physiological data collection methods to assess quantifiably the biomedical effects of cognitive workload on performance with instruments that can measure what the researchers cannot observe. Electroencephalography (EEG) machines utilize recording leads and different frequency bands to measure the micro patterns of brain activity while under stimulation (Sammer 2006). The instrumentation of the EEG can link the quantity and the quality of the tasks performed by the subject to their brain activity. Sammer hypothesized the point of cognitive overload could be identified with the EEG when the number of tasks exceeds the subject's ability to allocate mental resources. The combination of the EEG machine and a heart-rate variability monitor are excellent tools for collecting data on the effect of mental workload on task completion. The strength of this method is the ability of the EEG to collect data on brain activity when conducting tasks and while at rest. Still, the instrumentation is cumbersome and requires trained

professionals to read and interpret the data. Due to our research team's lack of expertise and compressed timeline, we cannot implement this method during our research.

Another method to assess cognitive workload is using validated survey instruments or subjective methods. The NASA-Task Load Index (NASA-TLX) is a qualitative cognitive workload measurement technique that assesses mental demand, physical demand, temporal demand, own performance, effort, and frustration level (Chen et al. 2016). Other subjective tools, such Multiple Resource Questionnaire and Bedford Scale (NASA 2020), have been used effectively. Although these techniques can measure mental workload, the performance assessment is subject to the operator's bias. According to Rusnock and Borghetti (2016), because retrospective subjective measures of workload are based on the average workload levels across a task, there is a possibility that the subjective workload levels may completely miss the fact that a cognitive overload had even occurred during mission execution. This potential for erroneous workload assessment is especially applicable when the researcher is modeling tasks with significant portions of low cognitive workload, which could lower the overall workload average. By not correctly accounting for relatively short periods of high workload, researchers can conclude specific missions are safe when, in reality, the pilot's workload may have spiked to unacceptable levels, which remain hidden due to the subjective workload assessment (Rusnock and Borghetti 2016).

While these subjective tools have seen success, they differ from data collection tools, such as the EEG, in a critical way. Subjective tools such as NASA-TLX and the Bedford Scale rely solely on operator estimation of their workload after the fact and generally only provide one value for the task. The summation of their effort leaves one with a cumulative workload total for that task, which is more likely to represent the average operator workload throughout the mission with no means to address workload variability (Rusnock and Borghetti 2016). Although the NASA-TLX was purposefully designed for aeronautics, due to the subjectivity of the self-evaluation, difficulty in collecting valuable, quantifiable data, and requirement for access to pilots, the research team is not utilizing these tools to collect data.

Once the data collection is completed, the data must be evaluated to determine if there is a correlation between different variables. The most common statistical analysis tools used to analyze data collected during cognitive workload studies were statistical techniques such as regression models, analysis of variance, and data tables to effectively establish relationships between two variables. Galy and Melan (2015) applied various mathematical approaches over a few weeks to analyze the effects of cognitive workload on performance and to model the impact of different individual characteristics contributing to cognitive workload. In their experiment, Galy and Melan required the participating students to complete various cognitive tasks under four different experimental conditions designed to induce stress and mental workload. Once collected, the researchers applied a statistical approach to analyze the data, and they were able to link workload successfully to performance. They confirmed that workload affected participants differently based on their individual characteristics.

Park, Fox, and Lang (2007) also implemented various statistical methods to identify a quantifiable relationship between information density, task complexity, and secondary task completion times. They analyzed the data they collected during their experiments to create a regression model to determine the strength of the relationship between variables. They successfully linked reduced secondary task performance due to an increase in the volume of dense and complex independent variables. Although both studies successfully established a relationship with data, we cannot utilize either of their approaches models because of time constraints and lack of access to participants. Our research team will utilize qualitative and descriptive methods to analyze data collected during interviews and from our digital simulator.

Another common approach uncovered in our research was using predictive models or simulations to assess the effects of mental workload on performance. Researchers utilized human-in-the-loop (HITL) experimentation numerous times, which Andrews et al. (2020) categorized as an effective data collection. Researchers use this method to accumulate information by building multiple prototypes or simulators to test the effects of workload on system operators. During HITL experimentation, human operators' performance and cognitive workload data are collected as they conduct

a simulated mission. Although HITL experimentation is an effective data collection method and could be valuable to our research, the time and cost associated with the construction of the simulators and the human operators are beyond the scope of our study.

Wickens' multiple resource theory (MRT) is a psychological theory at our research's core. His theory is a four-dimensional multiple resource model that predicts the effects of dual-task interference on a person's performance when required to perform complex tasks simultaneously (2002). Wickens posits that an individual's performance is negatively impacted as external task demands for attention exceed their workload capacity. His theory provides insight into how the human brain receives information and how the different sections process the data once received. Wickens' MRT theory is incorporated in the IMPRINT digital modeling software the team uses to measure the cognitive resources of MEDEVAC pilots during flight. In the simulation, the pilots will be presented with multiple demands for attention that engage their senses of perception and demand varying degrees of cognitive resources. The model is a cost-effective means to collect large amounts of data for the team that would be difficult to obtain without significant expenditure of time and finical resources.

Another common approach was using predictive models to assess the effects of mental workload on performance. Researchers utilized human-in-the-loop (HITL) experimentation numerous times, which Andrews et al. (2020) categorized as an effective data collection. Researchers use this method to accumulate information by building multiple prototypes or simulators to test the effects of workload on system operators. During HITL experimentation, human operators' performance and cognitive workload data are collected as they conduct a simulated mission. Although HITL experimentation is an effective data collection method and could be valuable to our research, the time and cost associated with the construction of the simulators and the human operators are beyond the scope of our study.

Park, Fox, and Lang (2007) Secondary Task Reaction Times (STRT) experiment also tested the negative effects of secondary task completion when the participant is engaged in a complex primary task. Although their model does not require intense

resources, their results confirmed Wickens' MRT. Their simple experiment required participants to identify different colors or sounds and either press a button or turn a key as fast as possible. The study's findings were similar to previous experiments and validated that people only have limited cognitive resources to process and perform tasks. Their methods and findings are informative to our research effort, suggesting there is a potential negative impact on human performance when multiple senses are engaged simultaneously.

In recent years, researchers have turned to synthetic environment modeling, and computer simulations are growing in popularity due to their flexibility, low cost, and ability to conduct an unlimited number of iterations with minimal manpower and resources. Wickens and Alexander (2009) successfully utilized a synthetic vision display to demonstrate the negative effect of cognitive overload on a pilot's ability to allocate mental resources to secondary tasks when performing complex primary tasks. They successfully created an immersive 3D simulation that would trigger intense mental focus in the pilots participating in the experiment. The intense focus of the pilots would result in a state of cognitive tunneling where the pilots would continue to apply additional mental resources to the primary task to the detriment of all other tasks. The pilots would continue to expend mental resources until they depleted their reserve, resulting in task failure. In another study on pilots' cognitive workload, Abreu et al. (2022) used synthetic modeling and computer simulation to measure an operator's cognitive workload during the steady state of flight for a MEDEVAC mission. This software allowed Abreu et al. to quickly and efficiently run dozens of iterations of the model to obtain statistically relevant information on the operator's cognitive workload and where that workload exceeded the operator's cognitive ability to process data input. Based on Abreu et al.'s success and considering our research team has similar constraints with time and resources, we decided to follow a similar approach with modeling and simulation.

A key ingredient required for successful computer simulations that measure cognitive workload is the tasks or events that drive a person's cognitive workload. Carter et al. (2021) identified five factors that were the most influential contributors to the cognitive workload of pilots based on a survey they conducted of MEDEVAC pilots. The

survey results determined that the environment, task demand, mission, operational competency, and fatigue were the contributing factors, with the environment and task demand being the most taxing. In a follow-up round of surveys, Carter et al. asked the pilots if they believed the cognitive workload they experienced could be mitigated by incorporating automation, and the pilots agreed with the research team's recommendation. Abreu et al.'s (2022) research aimed to identify which tasks could be automated and to determine if automating these tasks would reduce the pilot's cognitive workload. They utilized the modeling software to positively correlate the reduction in pilot cognitive workload to automating tasks, confirming the feedback provided by the MEDEVAC pilots in the research by Carter et al. The combined work of Carter et al. and Abreu et al. demonstrates a sound research technique to identify workload drivers and then isolate specific tasks to investigate to ensure a research project is manageable and can provide specific and measurable recommendations.

B. AUTOMATION IN FUTURE VERTICAL LIFT

Automation in the FVL is necessary to perform well in a dynamic and technological battlespace. Overall, researchers agree that incorporating automation in aviation has both positive and negative impacts. This section investigates the relationships between the operators and the use of automation by addressing the automation benefits, exploring how automation is integrated into modern technologies, and the problems associated with automation.

With respect to benefits, automating tasks can result in lowered cognitive workload, as shown by Abreu et al. (2022), Pop and Michelson (2018), and Young and Stanton (2007). Furthermore, automation in military aviation has resulted in improved flight performance, expanding the tactical capability of the aircraft and pilots, and reducing pilot workload during peak periods (Parasuraman et al. 1992).

With the increased technological advancements in aircraft, there has been a rise in new human error concerns. Therefore, it is imperative to understand the levels of automation and the options for partially automating tasks and review how the aviation community has perceived an automated agent by researching human-computer

interaction. Parasuraman et al. (1992) explain in their study the various levels of automation and expound on applying automation to aviation systems on a continuum. At one end, the operator would have complete manual control where the pilot controls everything continuously, or the helicopter will crash. On the other end, many operator tasks are fully automated tasks and may no longer be visible to the pilot (Parasuraman et al. 1992). In between these two ends of the spectrum are other levels that would balance the disparity between fully manual and fully automated.

Parasuraman et al. (1992) discussed ten levels, including one called consent automation. Consent automation is where the machine queues up a command or function but then must request permission from the pilot to implement its command (Parasuraman et al. 1992). Although consent automation, equating to level six out of the ten levels of automation, would help mitigate cognitive workload, Parasuraman et al. proposed adaptive automation as a more flexible solution for a cockpit.

Parasuraman et al. contended that adaptive automation could become the bridge between fully manual and fully automated systems. Adaptive automation is when the computer and the human can "trade tasks at appropriate times, in such a way as to maximize human performance" (Abreu et al. 2022, 13). Based on the pilot's performance and the current operating environment, adaptive automation has the automated system in the cockpit choose the level of automation (Parasuraman et al. 1992). However, these researchers note that adaptive automation would only help reverse unfavorable effects if applied to static or fixed automation. Likewise, Parasuraman et al. explained that the lack of interaction the pilot would have under the controls poses additional risks. These researchers explained that the pilot's lack of situational awareness of the system would negatively impact performance and pilot acceptance of automation.

Since the 1960s, technology has vastly improved in helicopters through studying human-computer interaction. In their studies, Lee and See stated that the goal of human-computer interaction in the system is to provide superior performance that could not be achieved alone by the human or the computer independently. Accordingly, automated agents in software and robotic settings have been designed and included to augment the pilots' interactions with the aircraft. An agent is "automation or another person that

actively interacts with the environment on behalf of the person" (Lee and See 2004, 51). Stone et al. (2010) determined that success would be found in creating an ad hoc agent. These researchers suggest that a competent automated agent must be developed with adjustable settings to interact with multiple types of teammates in unfamiliar environments to cooperate effectually. Considerations for expanding interactions include communication, mobility, and sensing capabilities (Stone et al. 2010). These researchers are confident that creating ad hoc autonomous systems will create flexibility to benefit collaboration between humans and computers without pre-coordination.

In addition, Lee and See endeavored to link the relationships humans have with other humans and how that translates to how humans depend on computers, trying to gain a better insight into perception and cognitive information processing. However, using human-human relationships as an analog may be challenging because automation lacks intention, loyalty, and benevolence (Lee and See 2004). Nevertheless, these researchers did find that integrating an automated agent into a system has appealed to the user's level of trust when computer etiquette, or personification of technology, has been incorporated into the system (Lee and See 2004). Furthermore, studies have shown that automation with good etiquette versus non-personified automation is favored by users and enhances operator performance, satisfaction, and trust (Lee and See 2004; Guyton 2021).

There are multiple concerns with trust, such as human error and the influence of out-of-the-loop communication caused by automation. Human error impacts trust because it leads to unintended consequences such as changes in behavior and the transfer of cognitive workload. Trust is also affected by out-of-the-loop communication, which leads to the degradation of manual skills and situational awareness (Endsley and Kiris 1995). Findings have shown that automation reduces workload while performing some tasks but increases it during others (Pop and Michelson 2018). Consequently, automation may alter behavior, often in unintended ways, leading to different inadvertent demands (Parasuraman et al. 2000). Parasuraman et al. explained that these behaviors occur because automation only partially replaces human activity.

Automation also can shift roles and responsibilities among the crew members in an aircraft. For example, in MacIsaac's (1998) research, she aimed to link

communication to crew performance. She noted a considerable decrease in communication interactions when communication tasks were automated. Conversely, when aviation tasks were automated, she said there appeared to be a shift of the same cognitive workload from one crew member to another instead of an overall decrease in workload. As a result, the crew's behaviors changed depending on the task demand. Understanding the difference between transferring and decreasing the cognitive workload will help our research team to understand the differences in the cognitive workload values between the operators when automation is introduced into the model.

Several researchers studied the impacts of out-of-the-loop communication. The out-of-the-loop (OOTL) phenomenon is a consequence of automation when there is a lack of cognizance of what the automation is accomplishing (Gouraud et al. 2017). In addition, researchers have verified that issues arise when the automation identified could cause the pilot to lose situational awareness and possibly even suffer manual skills degradation due to no longer performing that task (Endsley and Kiris 1995; Parasuraman et al. 1992).

According to a report from the National Transportation Safety Board (1994), a pilot crashed an aircraft just short of the runway because he confided too heavily on the autopilot during a night-time mission. The operator, professing little confidence in his manual control skills, failed to control the aircraft's speed during the final approach, which resulted in the accident (National Transportation Safety Board [NTSB] 1994). Loss of skills results from a lack of manual control training to help close the loop in an emergency and to correct errors when they occur. Manual control training will also help the pilot better understand the critical cues from the automation to get the pilot's attention and generate practical action.

In her research, MacIsaac (1998) elaborated that some unrealized errors, such as a lack of situational awareness of the flight control system modes (e.g., altitude, speed) and programming failures, have increased aviation accidents. Therefore, to minimize the likelihood of human error through a lack of situational awareness while operating the system, the user must be an active participant in the decision-making loop while maintaining a reduced cognitive workload.

Situational awareness can be improved through automation using monitors instead of manual controllers for a flight task because the displays offer a more comprehensive range of cues (Endsley and Kiris 1995). Their study showed that monitors help to distribute the users' attention better by integrating communication with the operators. However, monitors will only help prevent human error if there is not excessive information presented on the displays (Endsley and Kiris 1995). In addition, proper system feedback under automation is required for communicating problems as they arise to ensure the pilots respond as necessary.

Trust strongly influences the use of automation, mainly because users do not rely on it appropriately. There is a distinct relationship between reliance and trust and a user's experiences with automation and the criticality of the task. Humans know automation is not always perfect. Therefore, trust is a central issue with automation and has been seen since the introduction of the automated driving car (Parasuraman et al. 1992). Reliance and trust have all been shown to affect performance. Lee and See (2004) studied appropriate levels of reliance on automation and how trust is impacted. More experienced pilots have been shown to rely less on automation because they rely more heavily upon their abilities, even though automation would match the same level of performance (Lee and See 2004).

Overall, Lee and See found that trust varies based on individual differences. According to Stone et al. (2010), "a good team player's best actions are likely to differ significantly depending on the characteristics of its teammates" (1505). According to Singh, Molloy, and Parasuraman (1993), "high-complacency individuals actually detected more automation failures in a constant reliability condition (53.4% compared with 18.7% for low-complacency individuals)" (57). This statement means that more trusting individuals will trust automation more appropriately and adjust to various situations where automation is integrated (Lee and See 2004).

Research has also shown that familiarity with a system lessens perceptions of complexity and increases a user's confidence in using the system. Endsley and Kiris (1995) found that the confidence levels of the operators using fully automated systems were surprisingly high. This result is mainly because these users had previous experience

working with automated agents. These researchers found that experience gave the operators problem-solving confidence without losing skill in reacting to a system breakdown. Without familiarity with automation, studies have shown that users will opt out of employing automation if they feel they do not have time to engage in its operation.

Negative experiences with automation have a more considerable impact on the operator than positive experiences (Lee and See 2004). Parasuraman et al. (1992) discussed that if the system is not entirely dependable while performing its automatic function, then the pilot will not trust the automation and will intentionally or unintentionally monitor the task while it is being automated to determine if human intervention is necessary. Additionally, Lee and See (2004) noted that when the interviewed individuals had ample confidence in their ability to operate a system, they preferred to assume manual control than trust the automation. Lee and See (2004) affirmed that cognitive workload would be adversely impacted if trust in automation is compromised. Depending on the criticality of the task, the pilot may not feel comfortable doing other tasks while monitoring the critical task being automated, which violates the reason for automation to free the pilot's attention to perform other tasks simultaneously. The authors contended that this is a paradox in that one may inadvertently increase a pilot's cognitive workload rather than lower it as intended with the introduction of automation because the pilots continuously monitor the automation (Parasuraman et al. 1992). The researchers' study shows that there is a belief among pilots that automation in the cockpit has increased rather than lowered their cognitive workload, which has led to some pilots holding negative views about automation.

Abreu et al. (2022) focused on automating non-essential communication tasks that helped drive pilot and copilot cognitive overload but would be more acceptable to automate. This research team successfully identified which communication tasks could be fully automated without human interaction and could help reduce the aircrew's workload during their mission. Following their results, this team made the following recommendations for automated communications: monitoring radio nets, adjusting volume and input channels, selecting channels, and sending Joint Variable Message Format (JVMF) messages. According to Abreu et al. (2022), "automating these tasks

lowered pilot cognitive overload by 55.9%, copilot cognitive overload by 14.4%, and overall cognitive overload by 28%. Furthermore, the crew workload difference between the two models is statistically significant," as verified by their quantitative analysis (xviii). Noting these conclusions on the automation impacts on the cognitive workload of pilots and copilots, our research team intends to incorporate automation through all phases of flight. However, our team intends to incorporate automated tasks, including aviation, navigation, and communication tasks, to analyze the relationship between a high-level model and a model with more granularity and compare the results.

Researchers have concluded that automation must be targeted and deliberate for it to be helpful in reducing operator cognitive workload (Abreu et al. 2022; Parasuraman et al. 1992). These findings in this section reinforce the importance of designing and building automation that would reduce the pilot's cognitive workload and create an environment where pilots are confident and comfortable with how the automation would enhance their abilities, not replace them.

Our research into automation and lessons learned from Abreu et al. led our team to research methods to introduce automation into cognitive workload measurements. The following section highlights the tool of choice our team discovered based on research into measuring cognitive workload and utilizing automated agents.

C. IMPROVED PERFORMANCE RESEARCH INTEGRATION TOOL

To model the cognitive workload of the aircrew throughout the flight, the research team explored the IMPRINT, which is a trade-off analysis tool that can help evaluate cognitive workload and test alternate system-crew function and task allocations, such as the introduction of automation (Sargent 2022). IMPRINT began as an Airforce concept paper in the 1970s and has gone through several iterations and updates to become the product seen today (Sargent 2022). During the three-day IMPRINT training seminar he provided in 2022 at Naval Postgraduate School (NPS), Robert Sargent identified the simulation of missions as a primary use for IMPRINT.

Sargent identified four key inputs for successfully simulating a mission, as well as the outputs that IMPRINT provides. The inputs, as defined by Sargent, are the

identification of all tasks that will be completed in the mission, the times those tasks will take, which operator is the primary and secondary executor of each task, and the estimated visual, auditory, cognitive, and psychomotor (VACP) workload score for each task (Sargent, 2022). These scores are shown in Figure 2 (Rusnock and Borghetti 2016) and are the subjective estimate of subject matter experts regarding their cognitive load during task execution. These inputs align with the data needs that Abreu et al. (2022) discussed in their capstone project as well as the needs our research team identified for the project. This information will largely come from the team's research into MEDEVAC doctrine as well as interviews with actual UH-60 MEDEVAC pilots.

A common data collection method for IMPRINT is data collected through interviews with subject matter experts, which are used to populate the VACP demands as well as the workload value for each task. The workload evaluation method that IMPRINT uses is based on Wickens' multiple resource theory, which utilizes the VACP dimensions (Pop and Michelson 2018). To obtain the VACP dimensions along with the estimated workload score for each task, Abreu et al. (2022) conducted interviews with several MEDEVAC pilots. Salvi (2001) also conducted interviews with pilots to obtain VACP values for how much impact they perceived their equipment had on the time required to complete assigned tasks. During the interviews conducted by Abreu et al. (2022), pilots were asked to assign one or more of the VACP dimensions to each task along with an associated workload value, as shown in Figure 2. Abreu et al. (2022) then aggregated these scores across the interview population to determine the average workload for each task and then input the tasks, dimensions, and workload values into IMPRINT.

An example from Abreu et al.'s methodology would be if a pilot assigns a task as completely visual, the pilot will then be asked to assign an estimated workload value to that task utilizing the values in the visual column in Figure 2. If the pilot believes the task is heavily taxing on the visual dimension, they may assign it a score of 5.9 or 7.0. These scores, when assigned to all the tasks, begin to tell the story of the cognitive workload experienced by a pilot during their missions. Salvi (2001) used a similar subjective method for collecting the necessary data. Salvi first generated a list of equipment that Army doctrine said aviators would wear in the various scenarios he was exploring. He

then generated a detailed questionnaire that was provided to several Army aviators regarding what equipment they wore and how much it impacted the time it took them to execute a task. With a list of equipment as well as an estimated impact on task performance completed, Salvi then moved from the qualitative to the quantitative realm and had the pilots assign a value that detailed how much each piece of clothing negatively impacted task execution, if at all. According to Salvi, this section of the questionnaire was vital for building the IMPRINT model. Salvi then input the tasks, the baseline time pilots said it took to complete each task, and the impact their additional gear had on the completion time of that task into IMPRINT.

Care must be taken by a research team to assign accurate values regarding workload across the VACP based on the experience level of the operator who provided the workload values (Rusnock and Borghetti 2016). This is why the team will follow Salvi and Abreu et al.'s methods and use either the mean or median values of the data provided by the subject matter experts based on the size of the standard deviation between their answers. An example provided by Rusnock and Borghetti is that of driving a car. A beginner driver may lack experience and therefore assign a higher cognitive workload value to certain tasks, to which experienced drivers may assign low workload values (2016). These variances must be taken into account by a research team when the workload values are programmed into IMPRINT.

Value	Descriptors
	VISUAL
0.0	No Visual Activity
1.0	Visually Register/Detect (detect occurrence of image)
3.0	Visually Inspect/Check (discrete inspection/static condition)
4.0	Visually Locate/Align (selective orientation)
4.4	Visually Track/Follow (maintain orientation)
5.0	Visually Discriminate (detect visual difference)
5.1	Visually Read (symbol)
6.0	Visually Scan/Search/Monitor (continuous/serial inspection, multiple conditions)
	AUDITORY
0.0	No Auditory Activity
1.0	Detect/Register Sound (detect occurrence of sound).
2.0	Orient to Sound (general orientation/attention)
3.0	Interpret Semantic Content (speech, simple, 1-2 words)
4.2	Orient to Sound (selective orientation/attention)
4.3	Verify Auditory Feedback (detect occurrence of anticipated sound)
6.0	Interpret Semantic Content (speech, complex, sentence)
6.6	Discriminate Sound Characteristics (detect auditory differences)
7.0	Interpret Sound Patterns (pulse rates, etc.)
	COGNITIVE
0.0	No Cognitive Activity
1.0	Automatic (simple association)
1.2	Alternative Selection
4.6	Evaluation/Judgment (consider single aspect)
5.0	Sign/Signal Recognition
5.3	Encoding/Decoding, Recall
6.8	Evaluation/Judgment (consider several aspects)
7.0	Estimation, Calculation, Conversion
	FINE MOTOR
0.0	No Fine Motor Activity
2.2	Discrete Actuation (button, toggle, trigger)
2.6	Continuous Adjustive (flight controls, sensor control)
4.6	Manipulative (tracking)
5.5	Discrete Adjustment (rotary, vertical thumbwheel, lever position)
6.5	Symbolic Production (writing)
7.0	Serial Discrete Manipulation (keyboard entries)
	GROSS MOTOR
0.0	No Gross Motor Activity
1.0	Walking on level terrain
2.0	Walking on uneven terrain
3.0	Jogging on level terrain
3.5	Heavy lifting
5.0	Jogging on uneven terrain
6.0	Complex climbing
	SPEECH
0.0	No speech activity
2.0	Simple (1-2 words)
4.0	Complex (Sentence)
	TACTILE
0.0	No tactile activity
1.0	Alerting
2.0	Simple discrimination
4.0	Complex symbolic information

Figure 2. The Visual Auditory Cognitive Psychomotor Workload Scale. Source: Rusnock and Borghetti (2016).

The output that IMPRINT provides to the research team is the operator's workload throughout the duration of the mission. This output will allow the team to examine the operator's workload at any point during mission execution to identify where workload exceeds the threshold value, which is key to the stated goals of the research team and will allow the team to isolate those instances and introduce automation to determine if it has any impact on the workload of the operator. The general flow of creating an IMPRINT model follows the steps identified by Mr. Sargent during his 2022 lecture at NPS:

1. Define objectives
2. Design the study
3. Collect input data
4. Define the mission(s)
5. Develop the task data
6. Debug the model
7. Run the study
8. Collect output data
9. Analyze the data
10. Present the results

These ten steps provide a general outline of the approach that our research team will take in pursuit of our objective, which is to determine whether there is a difference in workload values between a high and low granularity model as well as the impact of automation on an aircrew's cognitive workload during a MEDEVAC flight.

A benefit of using IMPRINT for our research is the fact that IMPRINT has been utilized numerous times by other researchers for the specific goal of identifying and studying cognitive workload on humans (Sargent 2022; Pond et al. 2012; Pop and Michelson 2018; Salvi 2001; Powers and Gacy 2018; and Rusnock and Borghetti 2016). Sargent (2022) identified a trade study conducted for the mounted combat system that

was designed to determine if a two- or three-person crew was a more optimal solution for the vehicle. The study utilized IMPRINT to measure the cognitive workload of two different crew scenarios. The results of the study found that the two-person crew experienced unacceptable levels of cognitive overload during the execution of their mission sets, including during tasks deemed mission critical, which resulted in the program shifting focus from the two-person crew to a three-person crew which had acceptable levels of workload (Sargent 2022). While the research our team intends to do is not necessarily focused on optimizing crew levels, it is focused on identifying cognitive workload levels throughout the duration of the mission to identify and potentially mitigate spikes in workload levels over the IMPRINT threshold of 60.

Exploring how tasks impact operator cognitive workload has been researched and is the key focus of our research team's project. Pond et al. (2012) examined the allocation of communication tasks to reduce mental workload in pilots of remotely piloted aircraft (RPA). This study is very similar to the one our research team intends to execute in that Pond et al. specifically examined how communication tasks impacted the cognitive workload of pilots. Pond et al., much like Abreu et al., found that high levels and frequency of communication tasks while flying the aircraft results in above-threshold levels of cognitive workload. Pond et al. focused their mitigation efforts on changing text communication to audio communication and determined there was a 9% reduction in the time that pilot's cognitive workload exceeded the threshold value when 40% of received communications were changed from a text event to an auditory event. Pond et al.'s method differs from Abreu et al. and the method our research team plans to utilize. Whereas Pond et al. focused on changing the VACP dimension the task received in, Abreu et al. (2022) focused on introducing automation to select communication tasks, removing the need for the aircrew to do anything. Automation of the communication tasks resulted in a 28% reduction in cognitive workload for the aircrew (Abreu et al. 2022). Our research team intends to examine multiple IMPRINT models to determine if there is a difference in workload values between models with high and low task granularity. Our team also intends to introduce select task automation in these models and

determine if there is a difference in the impact automation has on workload values when implemented at differing levels of granularity.

IMPRINT provides flexibility when researchers need to introduce different variables or tasks to determine what provides the most significant impact on cognitive workload. According to Pop and Michelson (2018), the true value of IMPRINT is its ability to quickly and efficiently model the same route or mission while changing variables to determine which variable might be the most impactful on cognitive workload. This ability was demonstrated by Abreu et al.'s research, where they created a baseline model in IMPRINT, which was copied multiple times as a separate model (Abreu et al. 2022). This allowed Abreu et al. to introduce and isolate different variables in the separate models, which allowed a detailed comparison between the models and variables to determine which had the most significant impact on cognitive workload. Pop and Michelson utilized this as well by creating multiple iterations of their baseline model and then introducing different variables to each one so that they could compare the Tesla autopilot against the Cadillac Super Cruise. Sargent (2022) also highlighted this approach in the mounted combat system study, where a baseline model was built with two crew members and compared against an exact replica model with three crew members to determine which configuration would be operationally viable.

Another benefit of IMPRINT is that it allows the researcher to model the effects of automation by assigning tasks that you wish to automate to a phantom operator that can be built into the model (Sargent 2022). Every task in IMPRINT must be assigned to an operator who will perform that task. However, after you create your baseline model where the aircrew performs the tasks to determine your baseline workload levels, you can create additional models as mentioned above and assign tasks to be performed by the phantom operator, which mimics the task being automated. You can then filter out any tasks performed by this operator in the data collected at the end of each simulation, providing you with an updated cognitive workload value for the mission that accounts for the automation of tasks. Abreu et al. utilized this technique with IMPRINT. Pop and Michelson (2018) also utilized the phantom operator to simulate the various driving tasks they wished to automate in IMPRINT to help compare workload values against their

established baseline, which was a manually driven vehicle. Our research team intends to take a different path whereby we utilize updated VACP values provided by the pilots during knowledge elicitation. The pilots will be asked to provide updated VACP values for tasks they identify as candidates for automation. The identified tasks will receive the updated VACP values to simulate the operator's new workload for the automated task.

Another example of IMPRINT providing flexibility with workload modeling was shown by Powers and Gacy (2018), who utilized IMPRINT to model operator staffing and workload considerations in an operations center. Powers and Gacy focused on operator cognitive workload to determine which shifts were adequately staffed and which were improperly staffed. To simulate different staffing levels, the authors utilized the IMPRINT function of task reassignment, which allows tasks to be reassigned to secondary operators when the primary operator's workload values would exceed the threshold of the model. While this approach does not replicate the automation of tasks, it reinforces the validity of IMPRINT to measure and track cognitive workload, which is vital to our research.

One of the main benefits of using IMPRINT is its ability to measure workload dynamically rather than statically, as other mathematical and subjective methods do (Rusnock and Borghetti 2016). IMPRINT was specifically designed to model mental workload predictions and allows for dynamic tracking of the cognitive workload of an operator across the entire duration of the mission (Rusnock and Borghetti 2016). According to the authors, IMPRINT allows for the examination of the continuous analytic workload profile (CAWP), which allows researchers to examine a greater amount of detail regarding the operator's workload values across the entire task duration when compared to subjective rating methods previously mentioned. This is a significant benefit of utilizing IMPRINT over other methods because IMPRINT will not only show if an operator's workload levels are excessive, just like subjective methods will, but it also shows exactly where their workload crossed the threshold and which tasks are being executed when that event occurs (Rusnock and Borghetti 2016). This level of detail will help the research team isolate which tasks drive an aircrew's cognitive workload the most, as well as which tasks, when automated, reduce their workload the most. Abreu et

al. utilized this function of IMPRINT successfully to identify the exact moment the cognitive workload of the aircrew exceeded the threshold value and for how long it was above the red line. Abreu et al. were then able to isolate the tasks occurring at those times and introduce automation to specific tasks and compared the results against each other to determine if it reduced cognitive load. IMPRINT shows the overall workload level for the crew as well as the continuous workload of the aircrew across the entire task, which allows researchers to detect if and when the workload levels cross the threshold (Rusnock and Borghetti 2016).

Ultimately, the intent of IMPRINT is to inform possible future design decisions for the Army's Future Vertical Lift Program regarding the introduction of automation and management of the aircrew's cognitive workload levels. According to Rusnock and Borghetti (2016), IMPRINT's ability to inform system design decisions is its primary strength and another reason why our research team chose to utilize IMPRINT for our efforts.

D. CONCLUSION

Cognitive workload and modeling provided us with a conceptual understanding of what it is, how it is measured, and how it can be modeled in various studies. Mathematical modeling provides various statistical methods to identify a relationship between information density, task complexity, and secondary task completion times. Predictive models take the combination of data from the EEG machine and a heart-rate variability monitor, which is an excellent tool for collecting data on the effect of mental workload on task completion. Synthetic environment modeling and computer simulations are growing in their popularity due to their flexibility, low cost, and ability to conduct an unlimited number of iterations with a minimal expenditure of manpower and resources.

For automation in the Future Vertical Lift program, we reviewed the history of automation, the different levels of automation and their corresponding definitions, and how different levels of automation affect the level of trust in the system. This research will help minimize the cognitive overload of aircrews during MEDEVAC flights by optimizing their workload throughout the duration of the flight by automating specific

tasks that allow them to focus on tasks that demand expert knowledge, competence, and assessment. The research will also provide the HSA-DM project team with recommendations for which tasks would be beneficial to automate.

Lastly, our team researched the IMPRINT model. The workload evaluation method that IMPRINT uses is based on the multiple resource theory, which utilizes the VACP dimensions. A benefit of using IMPRINT for our research is the fact that IMPRINT has been utilized numerous times by other researchers for the specific goal of identifying and studying the cognitive workload of humans. IMPRINT provides flexibility when researchers need to introduce different variables or tasks to determine what provides the most significant impact on cognitive workload. One of the main benefits of using IMPRINT is its ability to measure workload dynamically rather than statically, as other mathematical and subjective methods.

We will use IMPRINT to model and simulate how certain tasks can be distributed between the pilot, copilot, and an envisioned AI system to mitigate the MEDEVAC pilot workload during a MEDEVAC flight. We will explore the areas of human systems integration, the environment, aviation tasks, navigation, and communication functions that are the causes of excessive cognitive workload.

In conclusion, the team conducted an extensive literature review to help us address the following issues: the impact of cognitive workload with automation in a cognitive model; the difference in cognitive workload as a function of automation and granularity levels; whether cognitive workload differs if more granularity is incorporated into a cognitive model; the impact on cognitive workload if automation is introduced into a cognitive model; and where in a MEDEVAC mission can pilot and co-pilot cognitive workload be reduced by introducing automation to the system at a subtask level.

III. METHODOLOGY

A. OVERVIEW

The primary purpose of this capstone project is to introduce automation into a previously validated IMPRINT model and determine the effect of automation on the cognitive workload of a UH-60 MEDEVAC pilot and co-pilot during a MEDEVAC flight. The research team will also determine if there is a difference in workload values across a simulated MEDEVAC flight from the perspective of a high-level model and another model with the same high-level tasks but with more detailed sub-tasks included. The research team will determine if there is a significant difference in workload values between these two models, both with and without the introduction of automation. In this chapter, we discuss the participants of the survey, the equipment used during the research, the procedures utilized by the research team, such as task analysis and knowledge elicitation, and the methods used to analyze the results.

The research team's approach began with an IMPRINT model provided by Mr. Scott Scheff that simulates a UH-60 MEDEVAC flight with several different variables to include ambulance exchange point operations, forward armament and re-fueling point operations, and multiple patient pick-ups (Scott Scheff, e-mail to author, November 14, 2022). Scheff built this model by using input from experienced UH-60 pilots, which the research team further validated by conducting additional surveys of more pilots to expand the model's sample size. The team applied the results of the surveys and updated Scheff's model as necessary based on the responses from the surveys. The team then introduced automation of specific tasks to the models and analyzed the results to determine the impact of automation on the cognitive workload throughout the flight. The research team interpreted these results and recommended tasks for automation to the HSA-DM team.

The team modified the Scheff model four times to achieve the research goals. The purpose of these modifications was to provide additional detail to the Scheff model and

automate select tasks to ascertain the impact of automation on cognitive workload. Table 1 provides an overview of the four different models the research team employed for this project.

Table 1. Overview of the Evolution of the Various IMPRINT Models

Model Version	Model Traits
Scheff 1.0	Baseline IMPRINT model with updated phase times
Scheff 2.0	Baseline IMPRINT model with automation and updated phase times
Seeing Sounds 1.0	Scheff 1.0 with additional tasks added
Seeing Sounds 2.0	Seeing Sounds 1.0 with automation

B. PARTICIPANTS

The participants of this research project were UH-60 MEDEVAC pilots, who provided the detailed information required to conduct a task analysis and update the Scheff 1.0 model. The targeted pilots were predominantly from the Army. The team conducted task analysis and knowledge elicitation with these pilots to provide a more robust sample size of pilots whose expertise was used to contribute to the VACP values in the Scheff 1.0 model. The research team interviewed ten UH-60 MEDEVAC pilots with a broad range of experience in terms of flight hours and years. However, only five pilots responded in time for their results to be included in this research. Table 2 displays each pilot's total flight hours and years of flight experience.

Table 2. Pilot Demographics

Pilot	Years of Experience	Total Flight Hours	Duty Position	Number of MEDEVAC Flights
P1	9	1300	Aeromedical Evacuation Officer	100+
P2	8	1227	Detachment Commander	40
P3	16	2200	Brigade S3	450
P4	10	1103	NATOPS Officer	2
P5	10	2000	Aeromedical Evacuation Officer	30-40
Total	53	7,830		627+

C. EQUIPMENT

The IMPRINT simulation software, previously discussed in Chapter II, was the main tool utilized by the research team to conduct the cognitive workload analysis. The primary capabilities of IMPRINT used in Scheff's 1.0 model were task analysis, VACP values, and organic reports, which are exportable to Microsoft Excel to help analyze the workload scores of the aircrew.

The team also used CmapTools, which is software that facilitates concept mapping to help us better understand Scheff's 1.0 model. A concept map is a visual way of organizing and presenting an idea or concept as a diagram and helps show the relationships or links between items (Jankowska 2014). The research team developed concept maps based on the IMPRINT model to facilitate knowledge elicitation with UH-60 pilots rather than attempting to familiarize them with the IMPRINT tool and its interface.

D. PROCEDURES

In this section, the research team provides an overview of the collection of data from subject matter experts, the creation of concept maps to visualize tasks being conducted during flight, and the procedures used to validate the VACP values contained within the Scheff 1.0 model. The research team conducted additional rounds of knowledge elicitation to increase the sample size utilized in the model, updated the model as necessary, and analyzed the results, which are covered in Chapter IV of this report. The following sections present a more detailed view of the methodology of our team for this project.

1. IMPRINT Model Validation and Construction

The research team's methodology approach was similar to the technique employed by Powers and Gacy (2018) in their research, which involved using IMPRINT to model operator staffing and workload considerations in a mission operations center (MOC). Our team developed a phased approach to the methodology to ensure we applied a disciplined and systematic process to review the current model. The team conducted a task analysis and knowledge elicitation for data collection, updated the baseline IMPRINT model, and analyzed the results. The initial phase began by exploring the Scheff model one task at a time. The detailed review reinforced our understanding and familiarization with the model and facilitated the creation of the concept maps. The team had to ensure that all four key inputs required for successfully simulating a mission in IMPRINT were accounted for during task analysis. As detailed by Robert Sargent (2022), the four key inputs are identifying all pertinent tasks, the time those tasks will take to complete, which operator is the primary executor of that task, and the estimated VACP workload score for each task. Powers and Gacy (2018) applied the task reassignment method in IMPRINT, which automatically reassigns a task to a secondary operator if the primary operator's workload exceeds the threshold value of 60. Our team did not use the transfer method for the Seeing Sounds 1.0 or 2.0 models. Rather we calculated the workload of the operator for the mission based on the strict VACP values in Seeing Sounds 1.0, even if those values exceeded 60. For Seeing Sounds 2.0, the team utilized

updated VACP values provided by the pilots with the assumption that the task would be automated. This assumption allowed the team to simulate the task being performed by an automated agent while being monitored by the operator. The reduction of the VACP values allowed the team to conduct a direct comparison between operator workloads in Seeing Sounds 1.0 and 2.0.

The team started the analysis by exploring each task level of the Scheff 1.0 model. The model contains three levels of tasks. The research team designated the highest-level tasks in the model as level-1 tasks. The level-1 tasks are decomposed into more detailed level-2 tasks, which are further decomposed into level-3 or terminating tasks. The level-3 or terminating tasks is the final level of the Scheff 1.0 model and contains the VACP data the research team requires to generate the operator workload throughout the simulation. A decomposition of the hierarchy of task types within the IMPRINT model is depicted in Table 3. After validating our understanding of task tier hierarchy, the team began updating and building the models we used to support our research.

Table 3. Overview of Task Levels of the Scheff 1.0 Model

Task Type	**Task Color in IMPRINT model**	**Example**
Model Task	Purple	
Level-1	Blue	
Level-2	Gray	
Terminating Task	Cyan—Pilot Task Green—Co-pilot Task	
Automated Task	Black	

To facilitate our research efforts, the team utilized the original Scheff model as the baseline model and derived all four models from it. Below is a brief description of

each model and how it is different from the original baseline model. For ease of reading we have highlighted the paths taken in each model, where applicable, with a red box.

a. Scheff Model 1.0

The Scheff 1.0 model is based on the original baseline model provided by Mr. Scheff. The only difference between the baseline model and the 1.0 model is that the research team adjusted the duration of each phase in the Scheff 1.0 model to match the phase durations in the Seeing Sounds 1.0 model to facilitate comparison and analysis. Figure 3 depicts the Scheff 1.0 perform hovering flight task path.

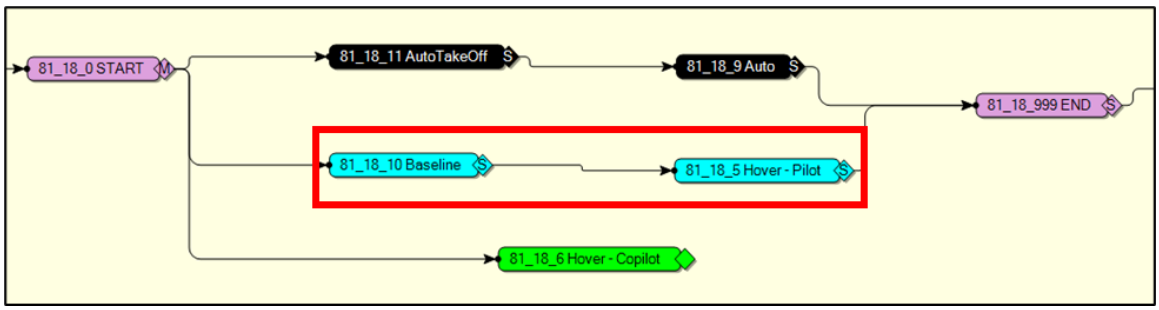

Figure 3. Scheff 1.0 Model Perform Hovering Flight Task

b. Scheff Model 2.0

The Scheff 2.0 model, as shown in Figure 4, is based on the original model with two key differences. In Scheff 2.0, the research team again adjusted the phase duration time to match the Seeing Sounds model duration times as well as activated the automated task path.

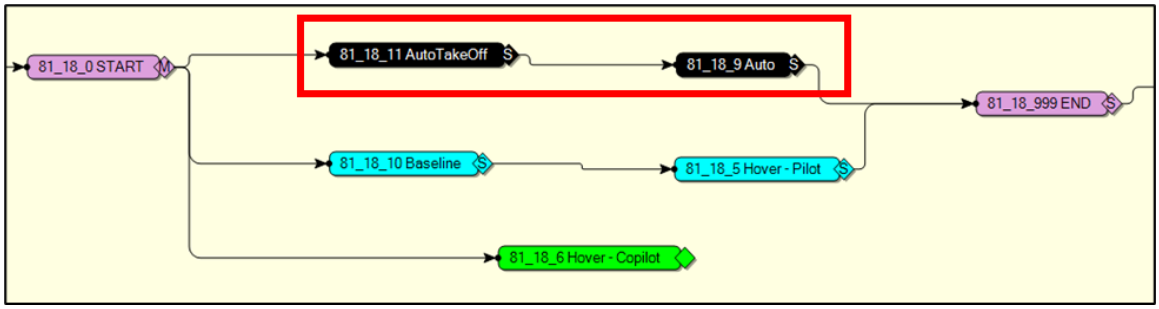

Figure 4. Scheff 2.0 Model Perform Hovering Flight Task with Automation

c. Seeing Sounds 1.0

Figure 5 depicts the Seeing Sounds 1.0 model, which is based on the baseline Scheff model as well. The major difference between the Scheff baseline and Seeing Sounds 1.0 is that the research team conducted a task analysis and added additional terminating tasks for more granularity.

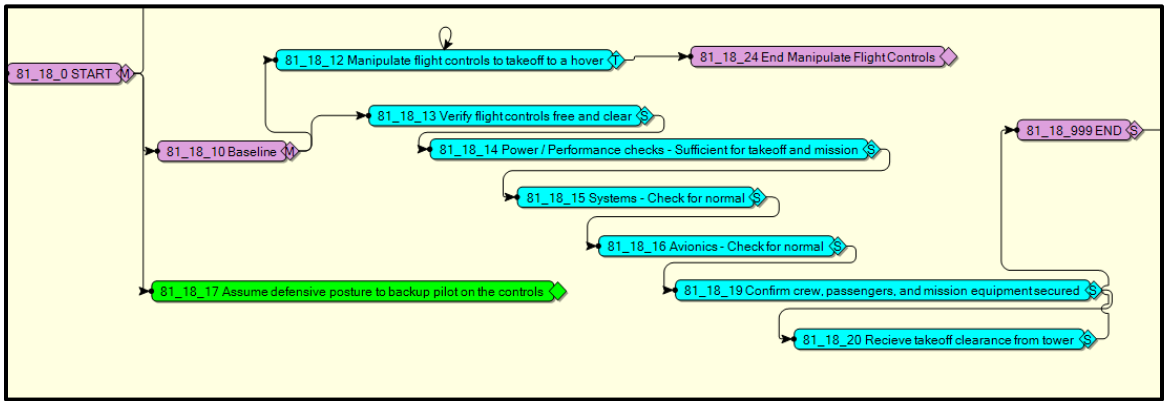

Figure 5. Seeing Sounds 1.0 Perform Hovering Flight Task with New Terminating Tasks

d. Seeing Sounds 2.0

Figure 6 depicts the Seeing Sounds 2.0 model, which is based on the Seeing Sounds 1.0 model but with the introduction of select task automation. For Seeing Sounds 2.0, the research team asked the pilots which tasks they deem beneficial for automation as

well as to provide updated VACP values for that automated task. The tasks identified for automation are colored black for ease of identification.

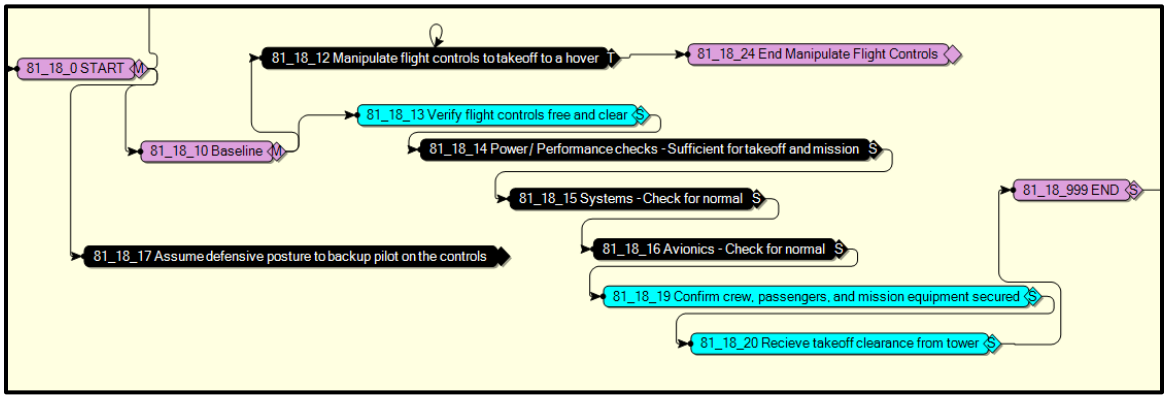

Figure 6. Seeing Sounds 2.0 Perform Hovering Flight Task with Automated Tasks

After reviewing each level of tasks in the IMPRINT model and identifying the terminating tasks, we transitioned to creating an associated concept map that mirrored the IMPRINT model.

e. 2.0 Concept Mapping

The concept maps generated by the research team with CmapTools software were designed to facilitate communication during our discussions with UH-60 pilots. The concept maps were used to simplify the complexity of the IMPRINT model into visual aids that are easier to understand and review than the model. A concept map is a visual way of organizing and presenting an idea or concept as a diagram and helps show the relationships or links between items (Jankowska 2014). Creating the concept maps, depicted in Figure 7, allowed the team to conduct our knowledge elicitation without providing time-consuming IMPRINT familiarization training to each interviewee.

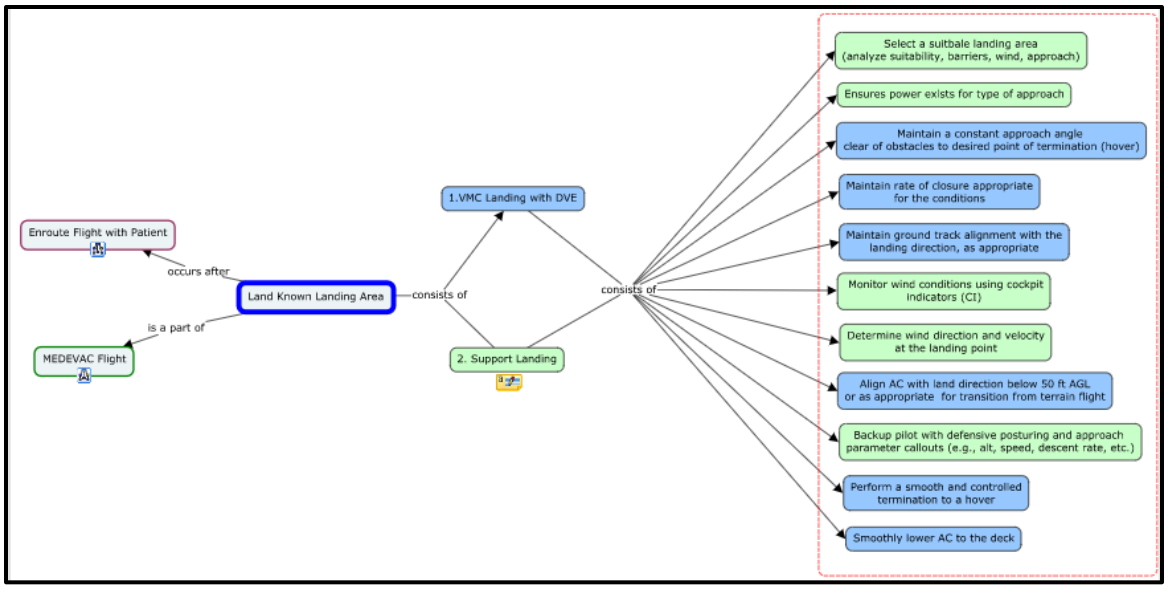

Figure 7. Land Known Landing Area Task Concept Map

Along with the concept map for each level-1 task, the research team created an Excel document listing each task, its associated level-3 tasks, and blank cells to input VACP values based on the knowledge elicitation with each pilot (Figure 8). This Excel document was then used to validate the original VACP values in the Scheff 1.0 model and update where applicable.

6. Land Known LZ															
High Level Task	Task	Operator	Terminating Task	Frequency	Length of Task	Standard Deviation	Visual	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tactile	Automation (Y/N)
Land Known LZ	VMC Landing & Support Landing	Co-Pilot	Select a suitable landing area (e.g., analyze suitability, barriers, wind, approach)	1	12 sec	3 sec	3	0	6.8	0	0	2	0	N
		Co-Pilot	Ensures power exists for type of approach	1	5 sec	3 sec	3	0	6.8	0	0	2	0	N
		Pilot	Maintain a constant approach angle clear of obstacles to desired point of termination (hover)	1	60 sec	10 sec	4.4	0	6.8	2.6	0	2	0	N
		Pilot	Maintain rate of closure appropriate for the conditions	1	60 sec	10 sec	4.4	0	6.8	2.6	0	2	0	N
		Pilot	Maintain ground track alignment with the landing direction, as appropriate	1	60 sec	10 sec	4.4	0	6.8	2.6	0	2	0	N
		Co-Pilot	Monitor wind conditions using cockpit indicators (CI)	2	5 sec	2 sec	3	0	1.2	0	0	0	0	N
		Co-Pilot	Determine wind direction and velocity at the landing point	1	10 sec	3 sec	4	0	1.2	0	0	0	0	N
		Pilot	Align AC with land direction below 50 ft AGL or as appropriate for transition from terrain flight	1	5 sec	2 sec	4	0	4.6	2.6	0	2	0	N
		Co-Pilot	Backup pilot with defensive posturing and approach parameter callouts (e.g., alt, speed, descent rate, etc.)	1	60 sec	10 sec	4.4	3	6.8	2.2	0	2	0	N
		Pilot	Perform a smooth and controlled termination to a hover	1	10 sec	5 sec	4	2	1	2.6	0	2	0	N
		Pilot	Smoothly lower AC to the landing pad	1	5 sec	2 sec	4.4	0	4.6	2.6	0	2	0	N

Figure 8. Task Analysis

After creating the concept maps and Excel documents to facilitate the team's knowledge elicitation activities, the team determined that it would benefit our stakeholders to further expand the Scheff 1.0 model by expanding the current level-3

terminating tasks with additional detail. Adding additional detailed tasks allowed the team to build a more robust model, which theoretically provides a more realistic workload value across the simulated flight compared to a higher-level model. The research team can also determine if a high-level IMPRINT model is more effective in determining cognitive workload values than a more detailed model. Attempting to expand the entire Scheff 1.0 model exceeded the scope of our efforts and could not be accomplished by the team due to the constrained timeframe for project completion. Given the team's constraints, the research team decided to narrow the focus of our effort and selected one distinct path within the Scheff 1.0 model to conduct a sufficiently robust task analysis.

2. Task Analysis

The team conducted a task analysis with subject matter experts to determine what additional tasks were needed to provide the additional tasks to the Scheff 1.0 model. An example of this can be seen in Figure 9, where the level-1 task is Land Known Landing Area. This task in the Scheff model consisted of two terminating tasks: VMC landing (Pilot) and Support Landing (Co-pilot), as shown in Figure 10. The team conducted a task analysis with subject matter experts to identify additional tasks beyond the original terminating tasks that would provide the research team with a more nuanced workload score across the entire flight, increasing the fidelity of the cognitive workload of pilots.

After consolidating the recommendations collected from the subject matter experts, the team expanded the depth of the Scheff 1.0 model by adding additional tasks. Figure 9 through Figure 11 provides an example from the Scheff 1.0 model of the different levels of tasks. Figure 9 shows the level-1 tasks in the model, and Figure 10 shows the terminating tasks, Figure 11 depicts the additional tasks added by the research team. Adding more terminating tasks also helped provide a more robust model to our sponsor and stakeholders at the project's conclusion. The tasks are color-coded by which operator is executing them. The pilot on the controls conducts tasks in cyan, tasks in green are performed by the pilot, not on the controls (or co-pilot), and tasks in black are those tasks chosen to be automated. The team determined it would be beneficial to assign

colors to tasks based on the operator who conducts them to represent more accurately how tasks are distributed during flight. Based on the Scheff 1.0 model, our research focus was exclusively on the pilot and co-pilot and did not examine any other members inside the UH-60 during the MEDEVAC flight.

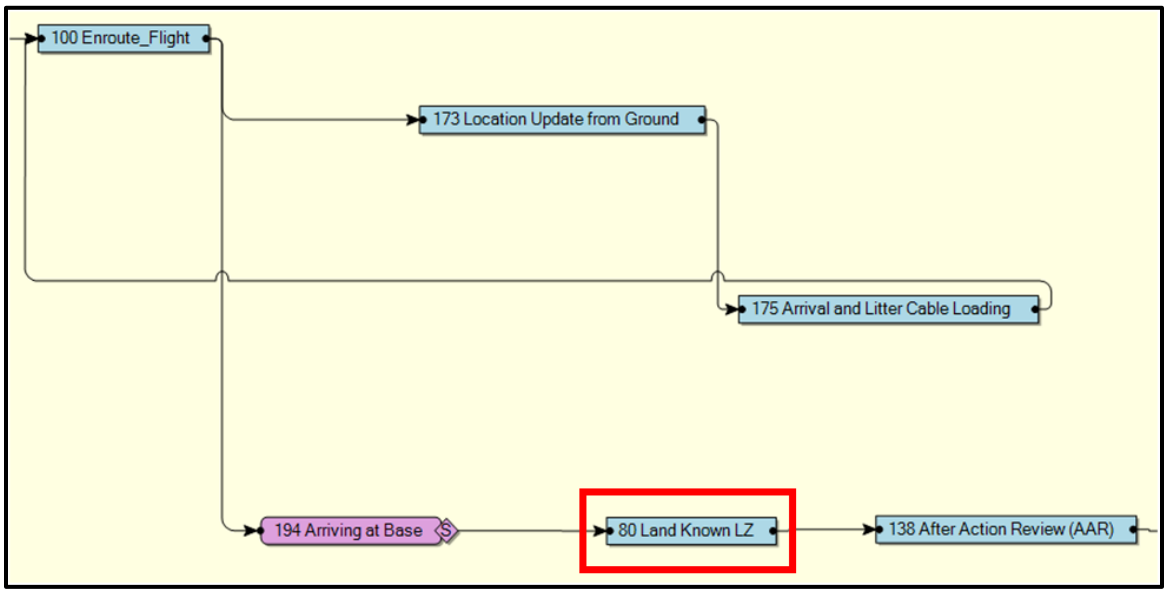

Figure 9. Level-1 IMPRINT Task Land Known Landing Area

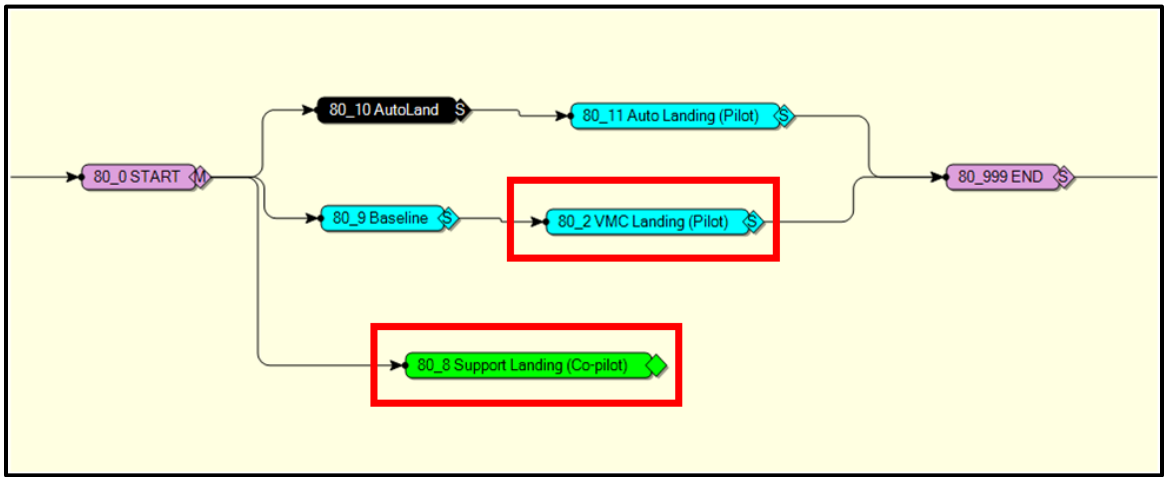

Figure 10. Level-2 Land Known Landing Area Scheff 1.0 Terminating Tasks

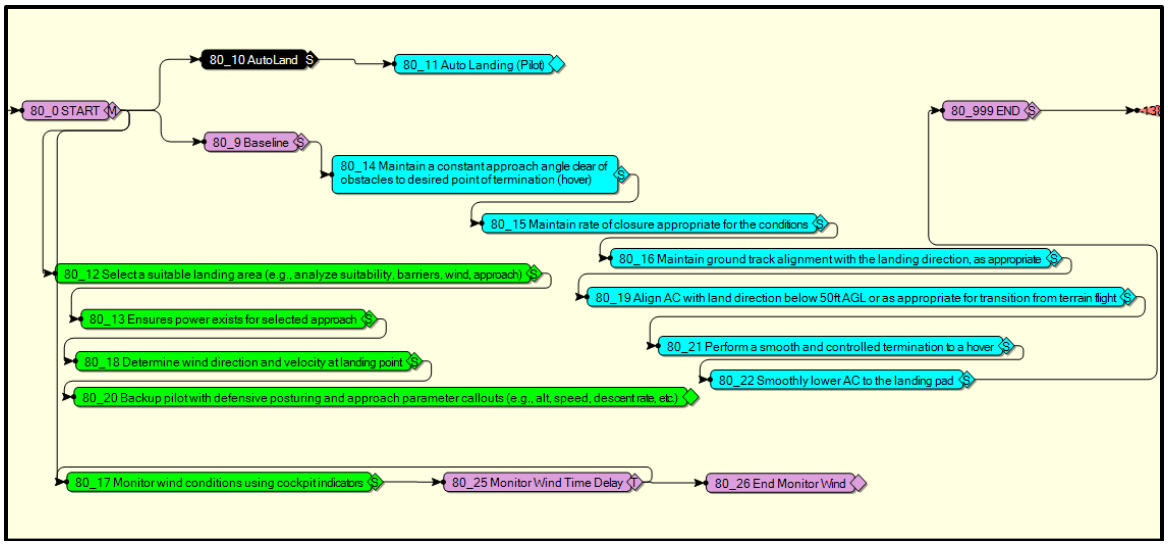

Figure 11. Level-2 Land Known Landing Area Seeing Sounds 1.0 Terminating Tasks.

3. Knowledge Elicitation

The team's knowledge elicitation approach was similar to Carter et al. (2021) and Abreu et al. (2022) in that we developed a standardized flight scenario based on the Scheff model to describe the situation in which the pilots would be operating. This helped narrow the discussion's focus and remove variables and unknowns that could detract from the pilots providing accurate VACP values.

Based on our research of Carter et al.'s and Abreu et al.'s knowledge elicitation methods, our team developed an initial demographic questionnaire, as shown in Appendix B, to determine the experience level of each pilot. The focus was the number of flight hours and flight years that each pilot had and what proportion of those flight hours were considered to be in a combat zone. The pilots were provided with this demographic questionnaire in advance of the actual knowledge elicitation phase to ensure they had time to fill it out and ask any questions they had of the process. In addition, each participant was provided with a brief overview of the project and a statement regarding the value of their participation in this research. The first step during the knowledge elicitation was to provide the pilots with an overview of the scenario for the mission in the IMPRINT model. The scenario, found in Appendix B, defined the environmental

conditions the pilots should consider while developing their VACP values as well as factors such as the expected threat level, the experience and fatigue levels of the aircrew, and what equipment the aircrew is using, such as night vision goggles. The team then walked the pilots through an explanation of the VACP factors and provided them with a brief example and description of what each value meant. Figure 12 shows the VACP values from Rusnock and Borghetti (2016) that were provided to each pilot for their reference during the process.

Value	Descriptors
	VISUAL
0.0	No Visual Activity
1.0	Visually Register/Detect (detect occurrence of image)
3.0	Visually Inspect/Check (discrete inspection/static condition)
4.0	Visually Locate/Align (selective orientation)
4.4	Visually Track/Follow (maintain orientation)
5.0	Visually Discriminate (detect visual difference)
5.1	Visually Read (symbol)
6.0	Visually Scan/Search/Monitor (continuous/serial inspection, multiple conditions)
	AUDITORY
0.0	No Auditory Activity
1.0	Detect/Register Sound (detect occurrence of sound).
2.0	Orient to Sound (general orientation/attention)
3.0	Interpret Semantic Content (speech, simple, 1-2 words)
4.2	Orient to Sound (selective orientation/attention)
4.3	Verify Auditory Feedback (detect occurrence of anticipated sound)
6.0	Interpret Semantic Content (speech, complex, sentence)
6.6	Discriminate Sound Characteristics (detect auditory differences)
7.0	Interpret Sound Patterns (pulse rates, etc.)
	COGNITIVE
0.0	No Cognitive Activity
1.0	Automatic (simple association)
1.2	Alternative Selection
4.6	Evaluation/Judgment (consider single aspect)
5.0	Sign/Signal Recognition
5.3	Encoding/Decoding, Recall
6.8	Evaluation/Judgment (consider several aspects)
7.0	Estimation, Calculation, Conversion
	FINE MOTOR
0.0	No Fine Motor Activity
2.2	Discrete Actuation (button, toggle, trigger)
2.6	Continuous Adjustive (flight controls, sensor control)
4.6	Manipulative (tracking)
5.5	Discrete Adjustment (rotary, vertical thumbwheel, lever position)
6.5	Symbolic Production (writing)
7.0	Serial Discrete Manipulation (keyboard entries)
	GROSS MOTOR
0.0	No Gross Motor Activity
1.0	Walking on level terrain
2.0	Walking on uneven terrain
3.0	Jogging on level terrain
3.5	Heavy lifting
5.0	Jogging on uneven terrain
6.0	Complex climbing
	SPEECH
0.0	No speech activity
2.0	Simple (1-2 words)
4.0	Complex (Sentence)
	TACTILE
0.0	No tactile activity
1.0	Alerting
2.0	Simple discrimination
4.0	Complex symbolic information

Figure 12. The Visual Auditory Cognitive Psychomotor Workload Scale. Source: Rusnock and Borghetti (2016).

Pilots were asked to provide details about each updated task, which included identifying which crew member performs the task, the frequency with which the task is

executed, the length of time spent performing the task, an estimated standard deviation for the time required to complete the task, and the estimated VACP workload value for that task. These values were recorded in our Excel document, as shown in Figure 8.

The pilots also were asked to provide recommendations on which tasks would be viable candidates for automation to help lower their cognitive workload during a MEDEVAC flight. These tasks were highlighted for further exploration regarding the feasibility of implementing automation.

4. Data Analysis from Interviews

The team examined the pilot responses to determine if there was concurrence regarding the VACP workload values assigned to the new terminating tasks. Where pilot responses varied, the research team adapted a mixture of mean, median, and mode to best represent the central value of their responses. The team analyzed the pilot's response regarding the duration of each task and the standard deviation in accomplishing each task. Where three or more pilots agreed on the duration and standard deviation, the research team utilized the mode value for the IMPRINT model. Where fewer than three pilots reached a consensus, we applied the median value for the duration and standard deviation. To calculate the combined VACP values for the IMPRINT model, the research team first calculated the standard deviation of the five responses received. The team utilized the median value, where the standard deviation between the responses was greater than one. The team utilized the mean value where the standard deviation between the responses was less than one.

5. Updating Scheff's IMPRINT Model

The research team collaborated with IMPRINT subject matter expert Robert Sargent to update the Scheff 1.0 model once all recorded values were calculated. We used his expertise to add the new terminating tasks with the updated task durations and VACP values to the model. Figure 13 shows the original Check AC Systems terminating tasks, while Figure 14 shows the updated Check AC Systems terminating tasks.

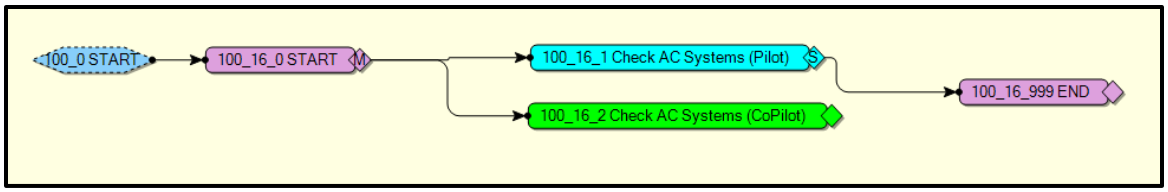

Figure 13. Check AC Systems Scheff 1.0 Terminating Tasks

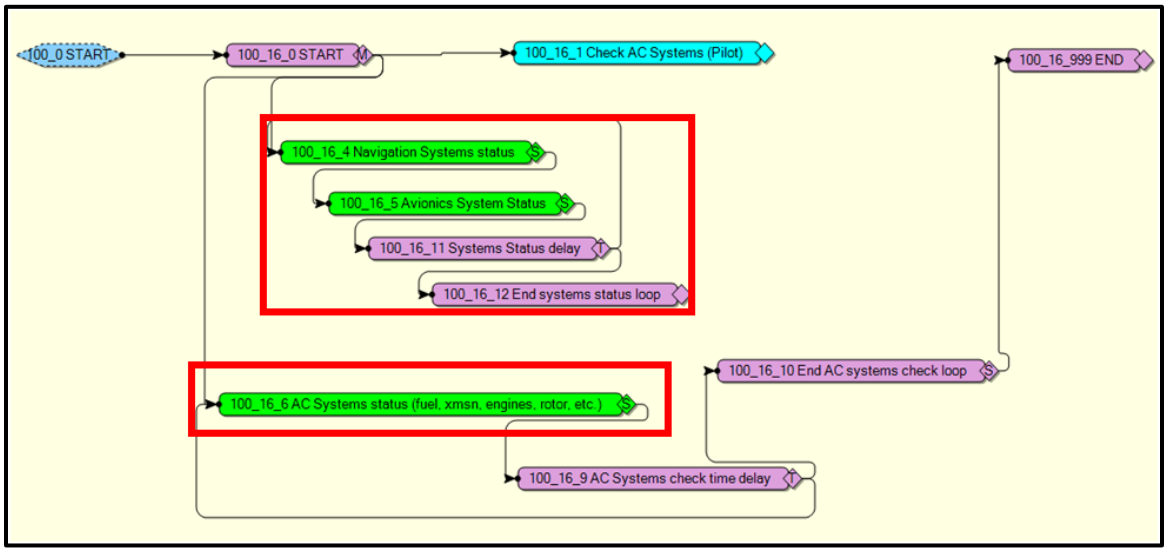

Figure 14. Check AC Systems Seeing Sounds 1.0 Terminating Tasks

The next step was to run each of the four models ten times. From these ten runs, we aggregated the data for each model to calculate the values necessary to compare the models against each other. The comparison of the Scheff 1.0 and the Seeing Sounds 1.0 models was intended to explore whether a significant difference in workload scores existed when utilizing a high-level or more detailed model. The comparison of the Scheff 2.0 and the Seeing Sounds 2.0 models was intended to explore whether a significant difference in workload values existed between a high and low-granularity model that included automation. The team analyzed the results with Robert Sargent to ensure that the Seeing Sounds 1.0 and 2.0 models were working correctly and that the workload values were accurately captured and portrayed. Any significant differences in workload levels were noted for further exploration.

6. Automating Tasks

To assist the team in determining and visualizing the cognitive workload demands on the pilot and co-pilot during a MEDEVAC flight, we utilized the built-in report function featured in IMPRINT that displays the operator workload report, operator workload graph, and operator overload report. The built-in tools in IMPRINT allowed the team to determine where in the flight the operator workload levels surpassed the threshold value of 60, how many times it breached and for what duration, as well as what tasks were being performed that contributed to the excessive workload. The operator workload graph visually depicts the operator workload over the duration of a MEDEVAC flight, as shown in Figure 15.

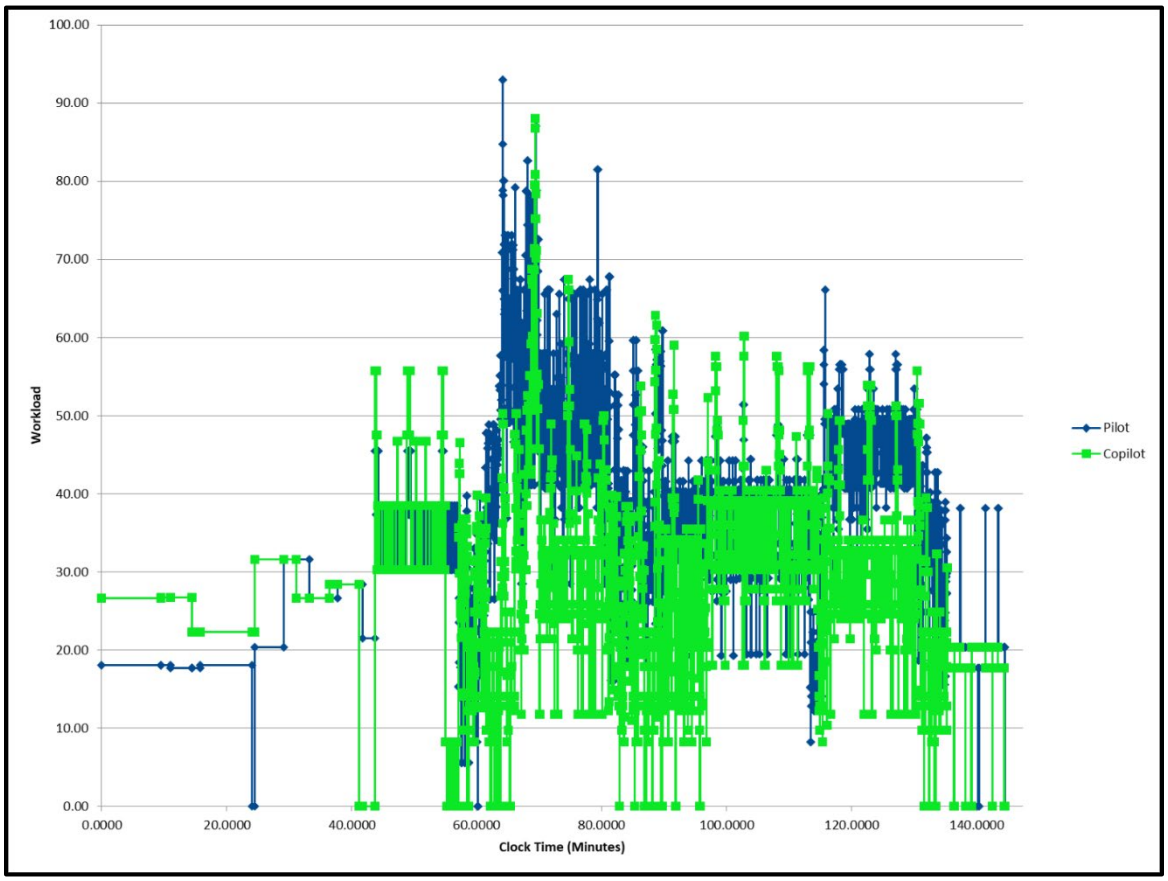

Figure 15. Seeing Sounds 1.0 Operator Workload Graph

The team identified the workload spikes over 60 and then cross-referenced the elevated levels against the operator workload report and overload report to determine which tasks were being performed at that time. After identifying the high-demand tasks, the team then referenced the input from our subject matter experts regarding which tasks they recommended to automate based on their experience. Any tasks identified as high workload drivers that correlated with the recommendations of the pilots were targeted as potential candidates for automation.

After establishing the baseline cognitive workload for the operators in the Scheff 1.0 and Seeing Sounds 1.0 models, the next step was to introduce automation into the models. The models with automation became Scheff 2.0 and Seeing Sounds 2.0, as shown in Figure 16 and Figure 17, respectively. The research team built the IMPRINT model to determine the impact of automation on the cognitive workload of the operators. After introducing automation to select tasks throughout the models, the team executed both models again to obtain updated pilot and co-pilot cognitive workload scores throughout the MEDEVAC flight.

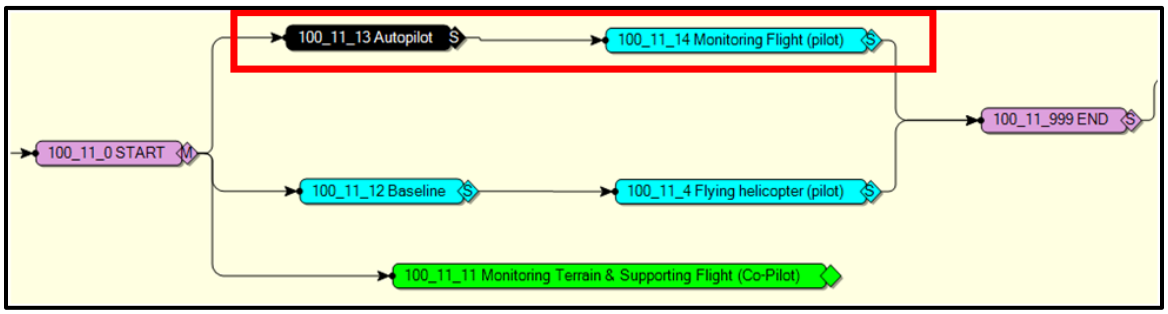

Figure 16. Scheff 2.0 with Automated Tasks Path Activated

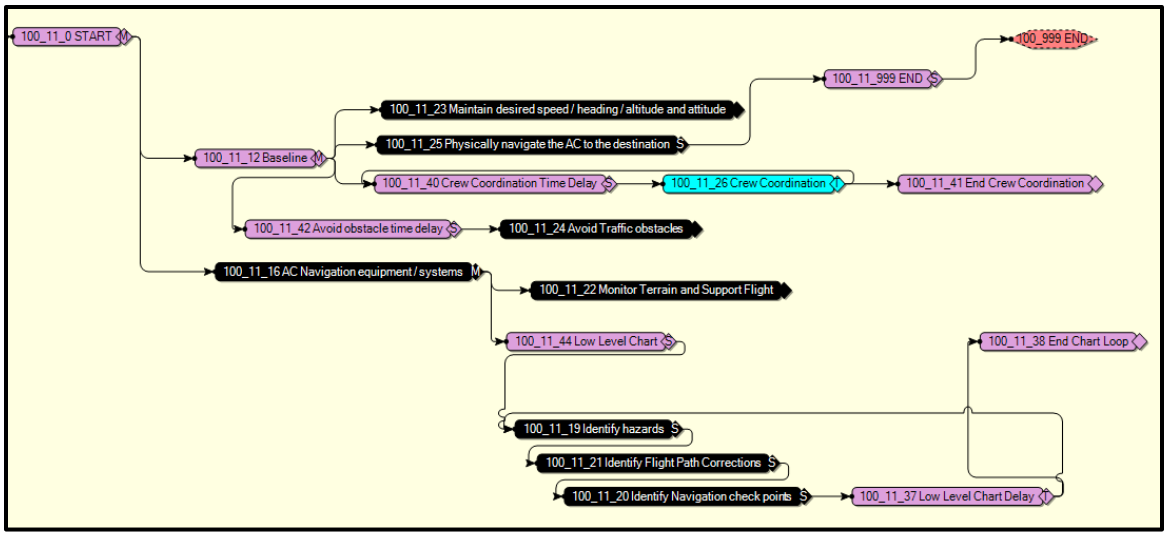

Figure 17. Seeing Sounds 2.0 with Automated Tasks in Black

Introducing automation into IMPRINT was relatively straightforward once the tasks were built into the Seeing Sounds 1.0 model. The research team replaced the original VACP values with new ones the pilots provided, assuming that the task was automated. The new values simulated the estimated workload that each operator would experience while the automated task was being executed. The updated workload values were predominantly lower than their original counterparts. However, the workload never went to zero because each operator would need to passively monitor task execution and completion at a minimum. The difference in workload values between Seeing Sounds 1.0 and 2.0 allowed the team to compare the differences in workload scores between a MEDEVAC flight with no automation and a MEDEVAC flight with automation which allowed us to provide targeted recommendations for task automation to HSA-DM. The results of our analysis and our recommendations are presented in Chapter IV.

IV. RESULTS

A. OVERVIEW

Using the methods described in Chapter III, our research team generated higher and lower granular cognitive workload models. We ran the Scott Scheff (SCH) 1.0 and Seeing Sound Team (SST) 1.0 baseline models for ten iterations. Identically, we ran the SCH 2.0 and SST 2.0 models with automation for ten iterations to conduct a comparative analysis of the four models. This chapter provides the key findings from the results and analysis of these models. Throughout this chapter, we provide the analysis and data to answer the research questions in Chapter I:

- Does cognitive workload differ if more granularity (subtasks) is incorporated into a cognitive model?
- What is the impact on cognitive workload if automation is introduced into a cognitive model?
- Is there a difference in cognitive workload as a function of automation and granularity levels?
- Where in a MEDEVAC mission, and to what extent, can pilot and co-pilot cognitive workload be reduced by introducing automation to the system at a subtask level?

By introducing subtasks into the cognitive workload model, we discovered an exponential increase in instances where the pilot's cognitive workload exceeded 60 and a 43% decrease in instances where the workload for the co-pilot exceeded 60. We also uncovered that automating subtasks did not change the average cognitive workload for the pilot in a substantive manner, but automating subtasks for the co-pilot did reduce the average cognitive workload by almost 9%. Additionally, subtasks provided a more detailed representation of the workload experienced by the pilot and the co-pilot than when it was represented in the lower granularity models. Next, the Visual Auditory Cognitive Psychomotor (VACP) scores for the pilots in the SST baseline model

compared to the SCH baseline model were significantly higher. This would suggest that the pilots interviewed in the SCH models believed that tasks associated with the pilot were significantly less demanding as compared to the pilots interviewed in the SST models. Lastly, the VACP scores for the co-pilots in the SST baseline model compared to the SCH baseline model were relatively similar.

B. COGNITIVE WORKLOAD: LOW GRANULARITY VERSUS HIGH GRANULARITY (SUBTASKS) COGNITIVE MODELS

This section will provide the data and analysis necessary to answer the first research question. We used the SCH workload model as our benchmark model that detailed a MEDEVAC flight based on pilot interviews with which we could compare the SST workload model. Table 4 and Table 5 portray the overall average and median workloads for the pilot and co-pilot in the SCH and SST 1.0 models. Lastly, Table 6 and Table 7 display the average VACP score for the pilot and co-pilot for the SCH and SST 1.0 models.

Workload occurrences over 60: The SST pilot 1.0 model has a much higher average (160.78) than the SCH pilot 1.0 model (0.4). Additionally, the SCH co-pilot 1.0 system has many more occurrences where the workload is over 60 (85) compared to the SST co-pilot 1.0 system (52.9).

Highest (MAX) Value: The SST pilot 1.0 model has a higher maximum workload value (87.38) than the SCH pilot 1.0 model (56.58). However, the SCH co-pilot 1.0 model has a higher average maximum value (98.82) compared to the SST co-pilot 1.0 model (78.51).

Overall average workload: The SST pilot 1.0 model has the highest overall average workload (39.39) compared to any of the other models. The SCH co-pilot 1.0 model has the second-highest overall average workload (32.26). The SCH pilot 1.0 model and the SST co-pilot 1.0 model had similar overall average workloads (31.66 and 30.08, respectively).

Table 4. Comparison of SCH Pilot and SCH Co-Pilot Workload, No Automation (10 Runs)

	SCH Pilot 1.0			SCH Co-Pilot 1.0		
	# over 60	**Highest (MAX) Value**	**Overall average workload**	**# over 60**	**Highest (MAX) Value**	**Overall average workload**
Average	0.4	56.58	31.66	85	98.82	32.26
Standard Deviation	0.84	3.14	0.41	27	11.39	1.16
Median	0	55.2	31.62	84	95.35	31.87
Total Count	4			851		

Table 5. Comparison of SST Pilot and SST Co-Pilot Workload, No Automation (10 Runs)

	SST Pilot 1.0			SST Co-Pilot 1.0		
	# over 60	**Highest (MAX) Value**	**Overall average workload**	**# over 60**	**Highest (MAX) Value**	**Overall average workload**
Average	160.78	87.38	39.39	52.9	78.51	30.08
Standard Deviation	9.73	2.72	0.57	28.38	7.1	1.03
Median	162	85.94	39.1	44.5	74.82	29.66
Total Count	1447			529		

Table 6. Comparison of VACP Scores of SCH 1.0 and SST 1.0, No Automation (10 Runs)

	Visual	**Auditory**	**Fine Motor**	**Cognitive**	**Gross Motor**	**Speech**	**Tactile**
SST Pilot 1.0	10.70	5.80	6.35	12.97	0.02	3.50	0.05
SCH Pilot 1.0	9.52	4.45	4.38	10.43	0.03	1.35	1.54
SST Co-Pilot 1.0	8.63	4.83	3.56	10.21	0.04	3.31	0.08
SCH Co-Pilot 1.0	8.79	5.05	4.31	10.77	0.05	3.31	0.58

Table 7. Comparison of Average Workload in SCH Cognitive Workload Model, Baseline versus Automated Mode

Overall Average Workload				
	SCH		SST	
	Pilot	Co-pilot	Pilot	Co-pilot
1.0	31.66	32.26	39.39	30.08
2.0	30.13	32.41	30.45	29.62

Based on the model scores after ten runs, the SST pilot 1.0 model data was higher than SCH Pilot 1.0 in every category except for Tactile (Figure 18). When examining the co-pilots, the SCH co-pilot 1.0 model has slightly higher modal scores in visual, auditory, fine motor, and tactile skills than the SST Co-Pilot 1.0. The cognitive and speech model scores are similar between the two models, while the gross motor modal score is almost negligible in both cases.

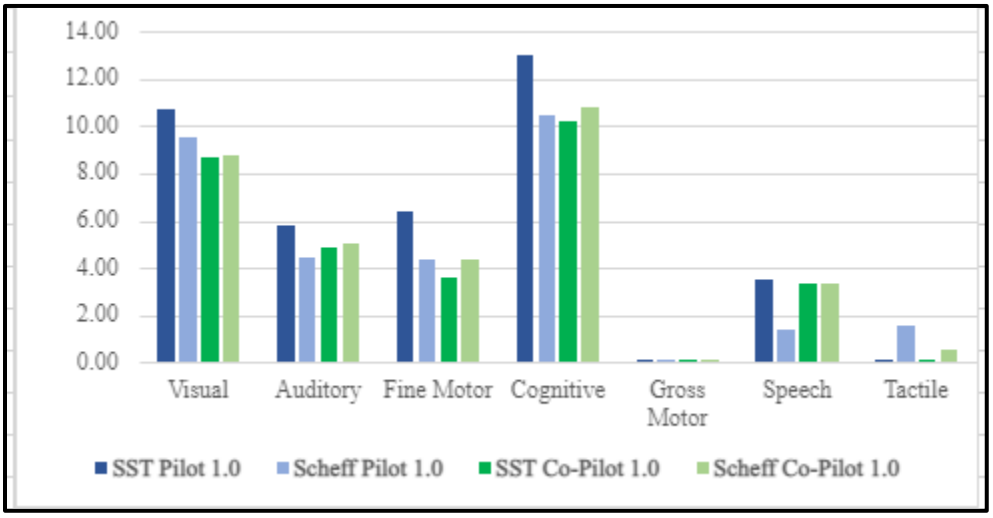

Figure 18. Graphic Comparison of Average VACP Scores of SCH 1.0 and SST 1.0, No Automation (10 Runs)

C. COGNITIVE WORKLOAD MODEL WITH AUTOMATION

We examined the overall average workload and VACP score for our second research question for the SCH and SST models. Table 7 portrays the overall average workloads for the pilot and co-pilot in the SCH 1.0 SST 1.0 and the SCH 2.0 SST 2.0 models. Table 8 compares the average, median, and total workload spike count in the SCH pilot 2.0 and co-pilot 2.0 models. Table 9 compares the average, median, and total workload spike count in the SST pilot 2.0 and co-pilot 2.0 models. Lastly, Table 10 and Figure 19 shows the VACP score between the SCH 2.0 and SST 2.0 models after ten model runs.

The SCH pilot 1.0 and SCH pilot 2.0 models have lower average cognitive workload scores than SST pilot 1.0 and SST pilot 2.0. The most notable difference is the 24% higher overall average workload in the SST pilot 1.0 model compared to the SCH pilot 1.0 model. The score difference between SCH co-pilot 1.0 and SST co-pilot 1.0 is relatively small, as with the overall average workload score between SCH co-pilot 2.0 and SST co-pilot 2.0.

Table 8. Comparison of SCH Pilot and SCH Co-Pilot Workload, with Automation (10 Runs)

	SCH Pilot 2.0			SCH Co-Pilot 2.0		
	# over 60	**Highest (MAX) Value**	**Overall average workload**	**# over 60**	**Highest (MAX) Value**	**Overall average workload**
Average	0.5	56.06	30.13	71	92.14	32.41
Standard Deviation	1.27	4.02	0.31	26	5.78	1.38
Median	0	55.3	30.16	64	90.6	32.31
Total Count	5			705		

Table 9. Comparison of SST Pilot and SST Co-Pilot Workload, with Automation (10 Runs)

	SST Pilot 2.0			SST Co-Pilot 2.0		
	# over 60	**Highest (MAX) Value**	**Overall average workload**	**# over 60**	**Highest (MAX) Value**	**Overall average workload**
Average	39.5	73.64	30.45	58	77.5	29.62
Standard Deviation	6.77	2.63	0.38	28.51	5.67	1.41
Median	39.5	72.3	30.28	48	76.57	29.71
Total Count	395			580		

The SCH co-pilot 2.0 model had the highest average number of occurrences where the workload score went above 60 after ten runs (71), followed by the SST co-pilot 2.0 model (58) and then by SST pilot 2.0 model (39.5). The SCH pilot 2.0 model had the lowest number of instances with a workload score over 60 (0.5). SCH co-pilot 2.0 model had the highest overall average workload score (32.41), followed by the SST pilot 2.0 model (30.45), then by SCH pilot 2.0 model (30.13), and lastly by SST co-pilot 2.0 model (29.65). The highest (maximum) workload score was recorded for the SCH co-pilot 2.0 model (92.14), followed by the SST pilot 1.0 model (73.64), then by SST pilot 2.0 model (73.64), and lastly by the SCH pilot 2.0 model (56.06).

Table 10. Comparison of VACP Scores of SCH 2.0 and SST 2.0, with Automation (10 Runs)

	Visual	**Auditory**	**Fine Motor**	**Cognitive**	**Gross Motor**	**Speech**	**Tactile**
SST Pilot 2.0	9.25	4.00	4.42	10.11	0.02	2.32	0.38
SCH Pilot 2.0	9.50	4.37	3.29	10.19	0.03	1.27	1.55
SST Co-Pilot 2.0	7.56	4.57	4.17	9.54	0.04	2.75	1.37
SCH Co-Pilot 2.0	8.87	5.04	4.28	10.79	0.05	3.32	0.60

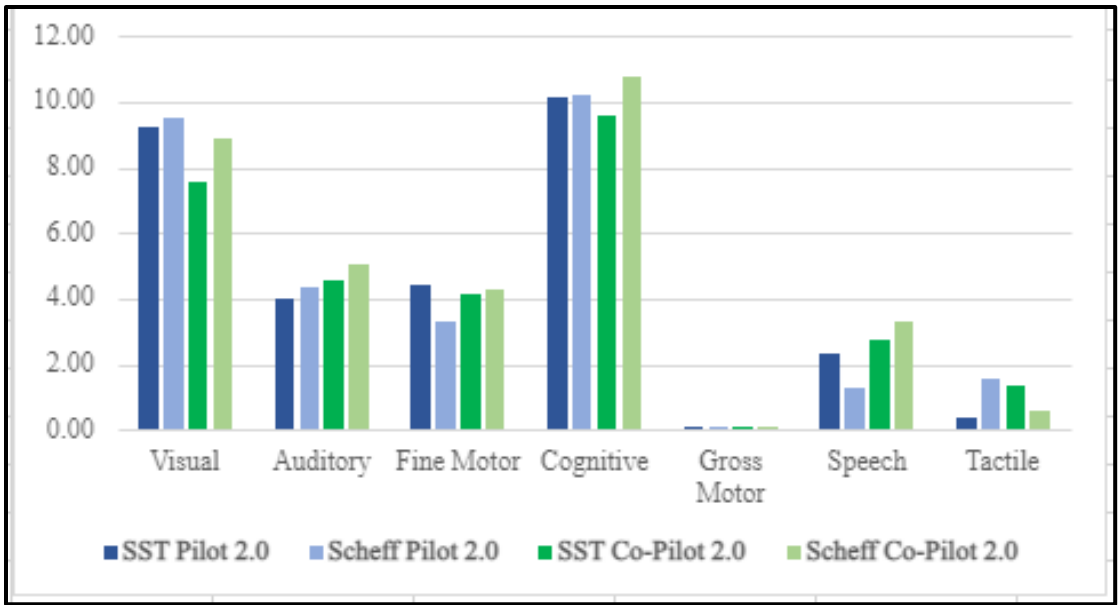

Figure 19. Graphic Comparison of VACP Scores of SCH 2.0 and SST 2.0, with Automation (10 Runs)

SCH pilot 2.0 scores slightly higher than SST pilot 2.0 in the visual, auditory, and tactile domains. SST pilot 2.0 scores slightly higher than SCH pilot 2.0 in the fine motor and speech domains. SCH co-pilot 2.0 scores higher than SST co-pilot 2.0 in all domains, except for auditory, where SST co-pilot 2.0 has a slightly higher score. Both models have the highest score in the cognitive domain, followed by the visual domain.

D. DIFFERENCE IN COGNITIVE WORKLOAD AS IT RELATES TO AUTOMATION AND GRANULARITY

For the next research question, the team compared the total and median workload spike occurrences for the SCH 1.0/2.0 and SST 1.0/2.0 models. Table 11 and Table 12 depict the total and median workload spike occurrences in the SCH and SST cognitive workload models. Lastly, Figure 20 graphically represents the VACP score between the SCH 1.0/2.0 and SST 1.0/2.0 models for comparative analysis.

Table 11. Comparison of Total and Median Workload Spike Occurrences in SCH Cognitive Workload Model, Baseline versus Automated Model

	Pilot		Co-pilot	
	Total	**Median**	**Total**	**Median**
SCH 1.0	4	0	851	84
SCH 2.0	5	0	705	64

Table 12. Comparison of Total and Median Workload Spike Occurrences in SST Cognitive Workload Model, Baseline versus Automated Model

	Pilot		**Co-pilot**	
	Total	**Median**	**Total**	**Median**
SST 1.0	1447	162	529	45
SST 2.0	395	40	580	48

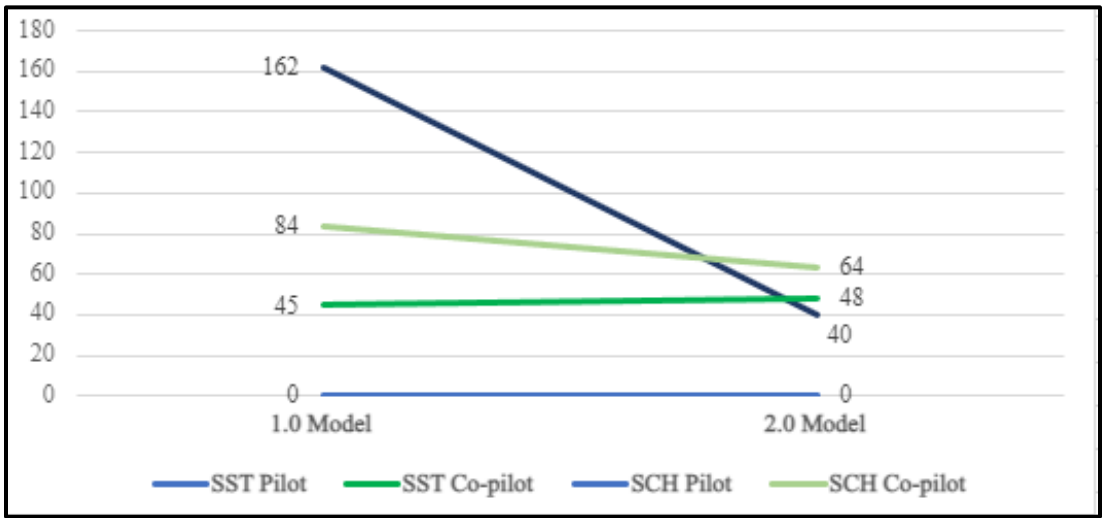

Figure 20. Graphic Comparison of Median Workload Spike Occurrences in the SCH and SST Cognitive Workload Model, Baseline versus Automated Model

There is a difference in the frequency of median workload spikes between the SCH and SST cognitive workload models. In the SCH model, there are no median workload spike occurrences for the pilot in both the 1.0 and 2.0 models, while for the co-pilot, there are 84 median workload spike occurrences in the 1.0 model and only 64 in the 2.0 model. In contrast, in the SST model, the median workload spike occurrences for the pilot are higher for the 1.0 model (162) compared to the 2.0 model (40), while for the co-pilot, the median workload spike occurrences are higher for the 2.0 model (48) compared to the 1.0 model (45).

E. COGNITIVE WORKLOAD REDUCTION WITH AUTOMATION TO THE SYSTEM

Lastly, for the fourth research question, the team examined the lowest and highest cognitive workload out of the ten runs for the SCH 1.0/2.0 and SST 1.0/2.0 models. Figure 21 and Figure 22 depict the lowest and highest workload graphs in the SCH 1.0 models. Figure 23 and Figure 24 depict the lowest and highest workload graphs in the SST 1.0 models. Figure 25 and Figure 26 depict the lowest and highest workload graphs in the SCH 2.0 models. Figure 27 and Figure 28 depict the lowest and highest workload graphs in the SST 2.0 models. Figure 29 depicts pilot tasks that exceed a cognitive workload of 60 in the SST 1.0 model for all ten runs. Figure 30 depicts co-pilot tasks that exceed a cognitive workload of 60 in the SST 1.0 model for all ten runs. Figure 31 depicts co-pilot tasks that exceed a cognitive workload of 60 in the SCH 1.0 model for all ten runs. Figure 32 depicts co-pilot tasks that exceed a cognitive workload of 60 in the SCH 2.0 model for all ten runs. Figure 33 depicts pilot tasks that exceed a cognitive workload of 60 in the SST 2.0 model for all ten runs. Figure 34 depicts co-pilot tasks that exceed a cognitive workload of 60 in the SST 2.0 model for all ten runs.

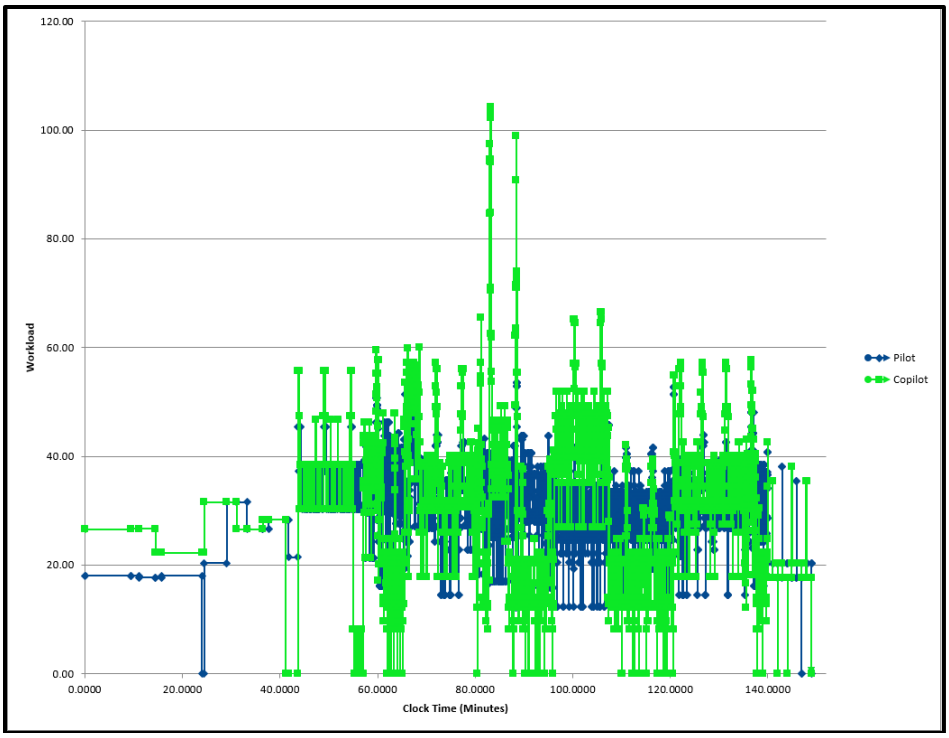

Figure 21. Lowest SCH Workload Graph 1.0, No Automation (Run #3 of 10)

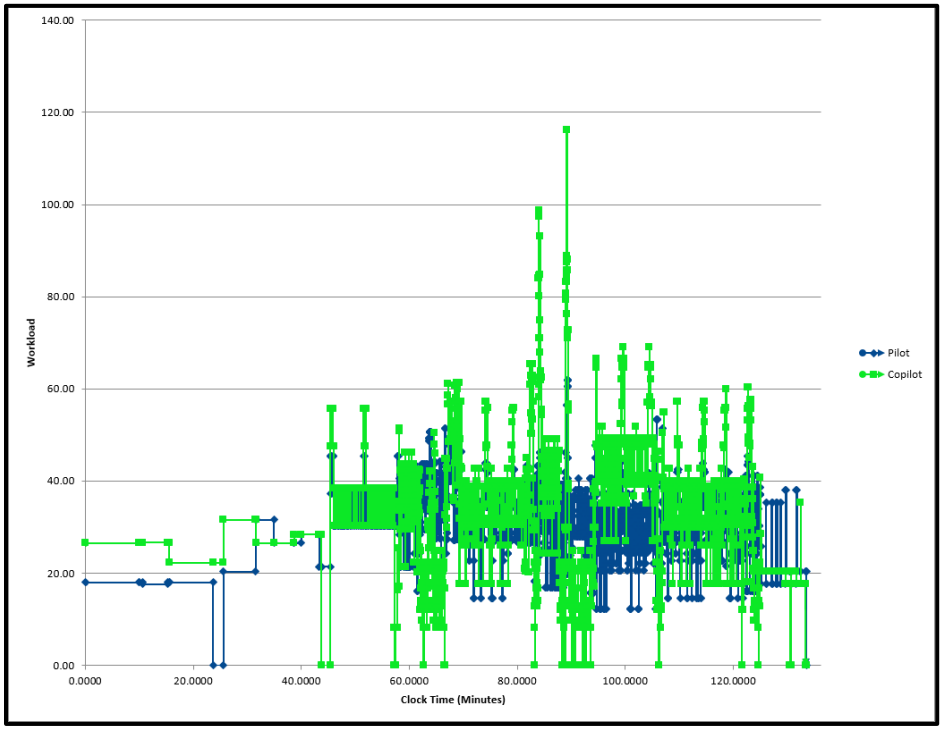

Figure 22. Highest SCH Workload Graph 1.0, No Automation (Run #4 of 10)

The co-pilot in the SCH 1.0 model has a significantly higher number of instances where their cognitive workload was over 60 compared to the pilot in the SCH 1.0 model. In runs three and four, the pilot in the SCH 1.0 model had 0 and 2 instances, respectively, while the co-pilot in the SCH 1.0 model had 40 and 82 instances, respectively. Additionally, the SCH 1.0 model co-pilot had higher maximum workload values in runs three and four compared to the pilot in the SCH 1.0 model. However, the average workload for the co-pilot in the SCH 1.0 model in runs three and four were lower than that of the pilot in the SCH 1.0 model.

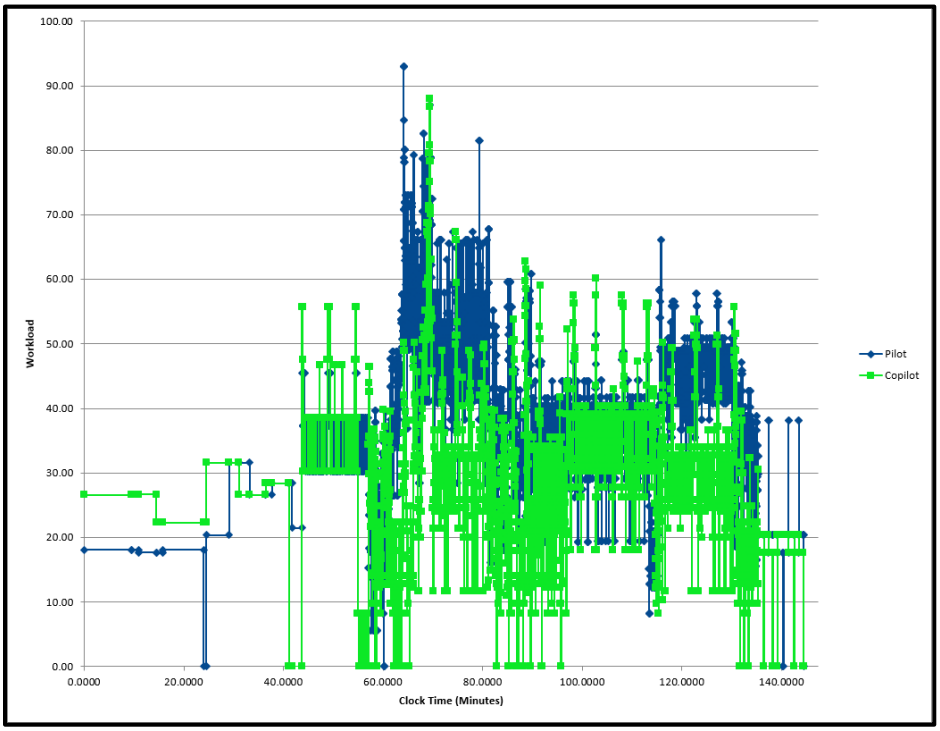

Figure 23. Lowest SST Workload Graph 1.0, No Automation (Run #3 of 10)

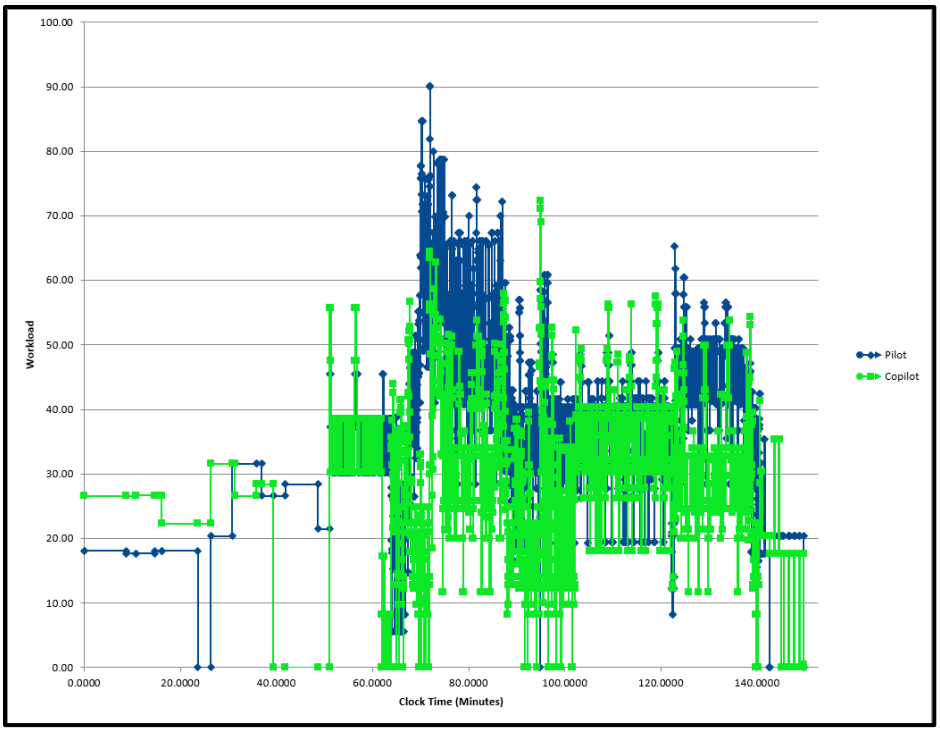

Figure 24. Highest SST Workload Graph 1.0, No Automation (Run #6 of 10)

On the other hand, the pilot's workload in the SST 1.0 model is higher than that of the co-pilot in both the run three and six scenarios. This higher workload is indicated by the higher overall average workload values for the pilot in both cases (39.71 and 40.37) compared to the co-pilot (28.94 and 29.71). Additionally, the pilot had more instances where their workload was over 60 compared to the co-pilot in both scenarios. The co-pilot also had a lower highest (MAX) workload value in both scenarios, which suggests that they did not experience as extreme workload peaks as the pilot did.

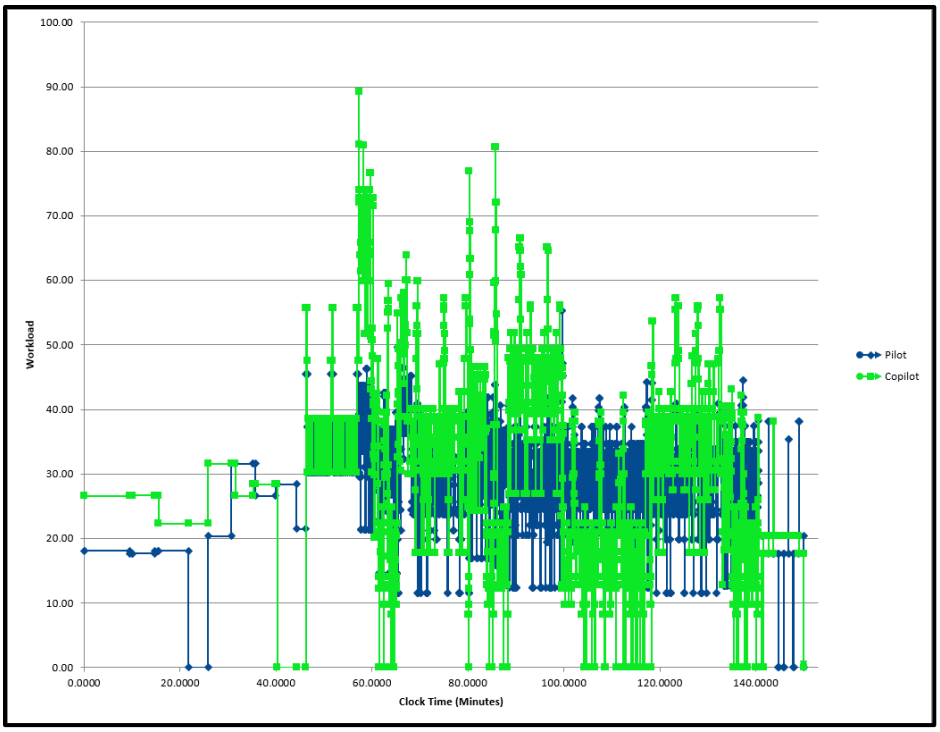

Figure 25. Lowest SCH Workload Graph 2.0, with Automation (Run #9 of 10)

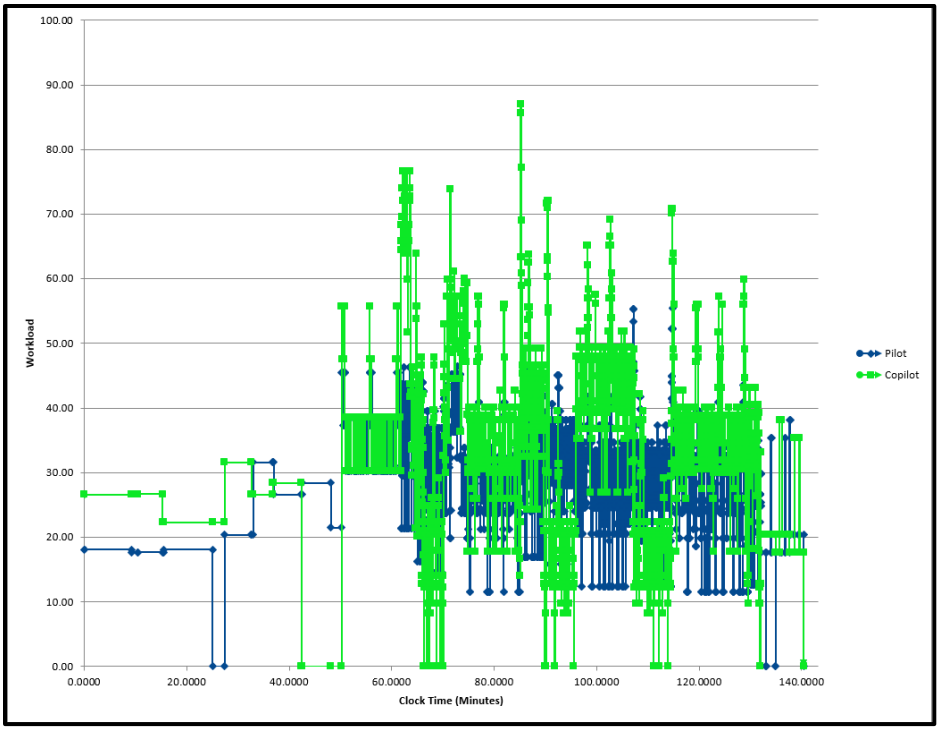

Figure 26. Highest SCH Workload Graph 2.0, with Automation (Run #10 of 10)

Comparing the pilot and co-pilot in the SCH 2.0 models, the pilot had a much lower workload than the co-pilot in both the run nine and ten scenarios. The pilot had no instances where their workload was over 60, whereas the co-pilot had over 100 instances in both scenarios. Additionally, the overall average workload for the pilot was lower in both cases (29.57 and 30.08) compared to the co-pilot (31.97 and 35.2). However, the co-pilot had a higher highest (MAX) workload value in both scenarios, which suggests that they experienced more extreme workload peaks at times. Nonetheless, the absence of instances where the pilot's workload was over 60 indicates that the pilot's workload was generally more manageable than that of the co-pilot.

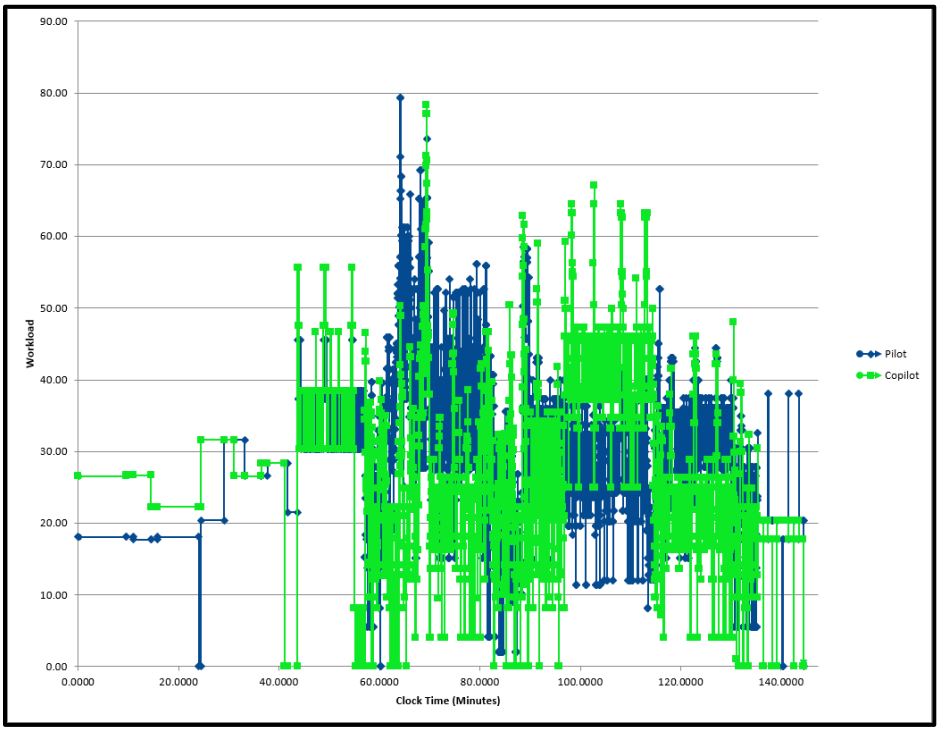

Figure 27. Lowest SST Workload Graph 2.0, with Automation (Run #3 of 10)

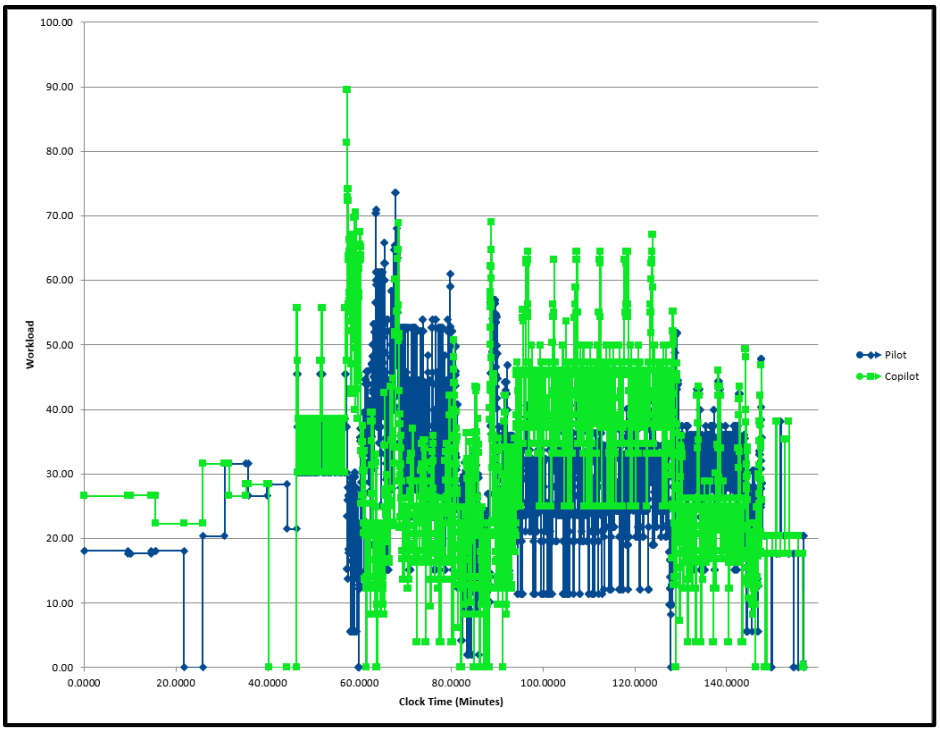

Figure 28. Highest SST Workload Graph 2.0, with Automation (Run #9 of 10)

Comparing the pilot and co-pilot in the SST 2.0 model, the pilot had a similar or slightly higher workload in both run three and nine scenarios compared to the co-pilot. In the run three scenario, the pilot had slightly more instances where their workload was over 60 (38 compared to 34 for the co-pilot), a slightly higher highest (MAX) workload value (79.3 compared to 78.36), and about a 10% higher overall average workload (30.67 compared to 27.67) than the co-pilot. In the run nine scenario, the pilot had slightly fewer instances where their workload was over 60 (43 compared to 99 for the co-pilot), a slightly lower highest (MAX) workload value (73.6 compared to 89.5), and a slightly lower overall average workload (30.09 compared to 32.28) than the co-pilot.

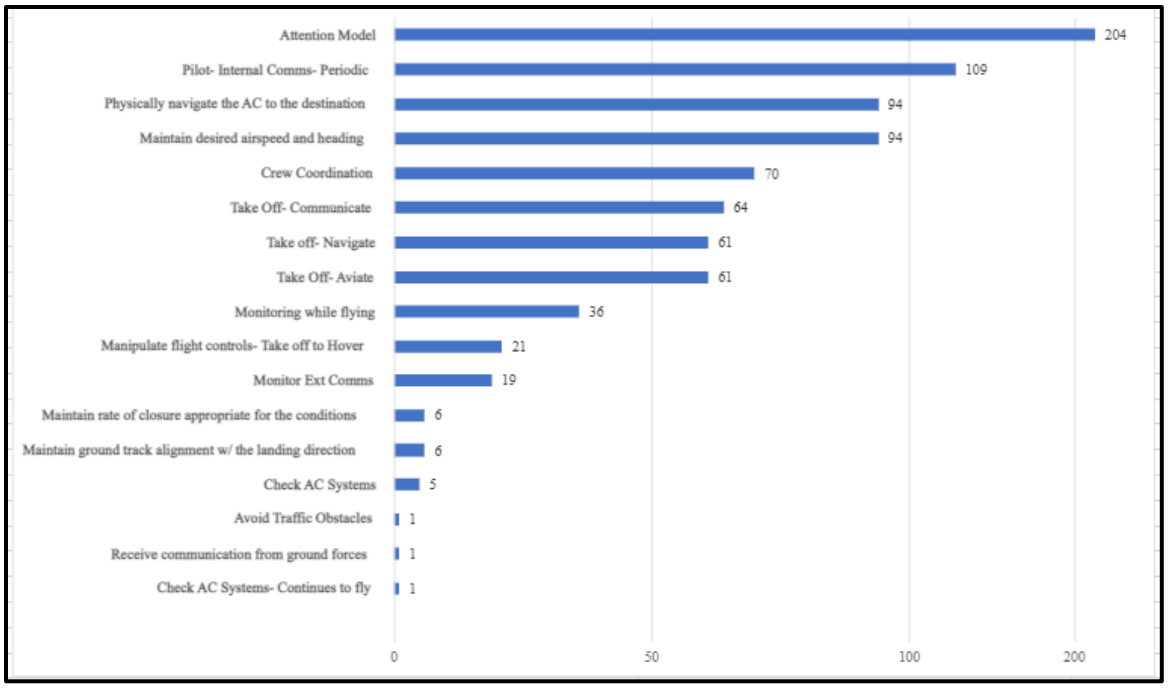

Figure 29. Pilot Task Over 60 SST 1.0 Model (Highest Run)

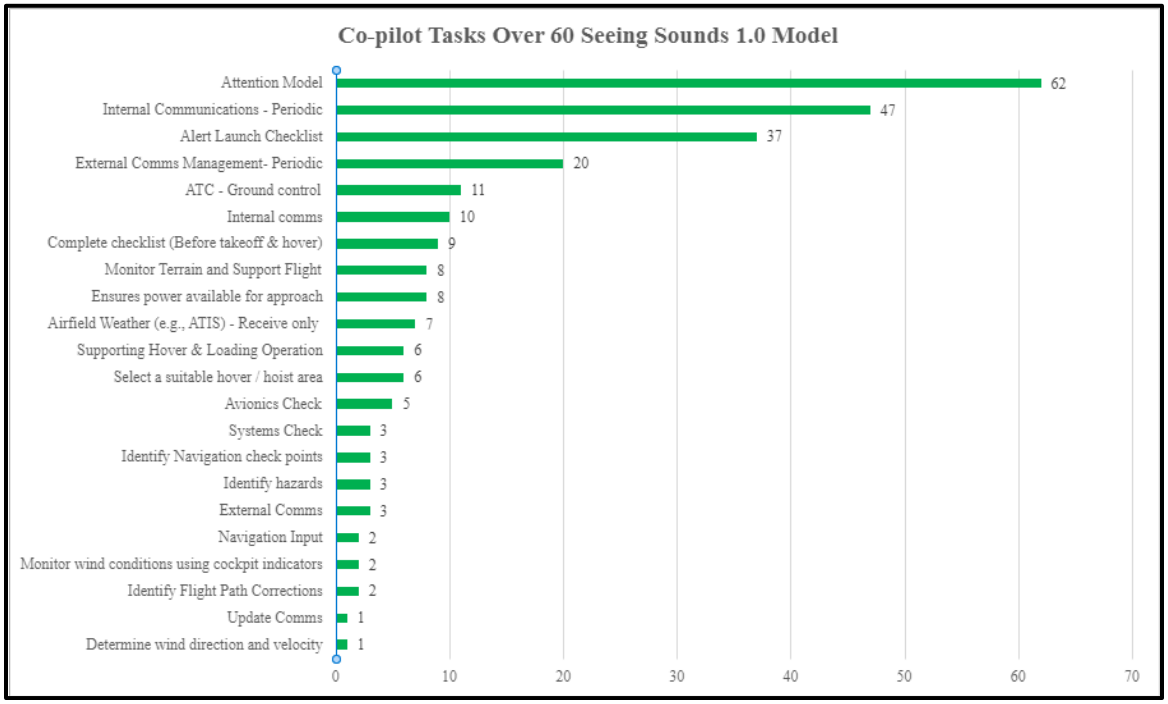

Figure 30. Co-pilot Task Over 60 SST 1.0 Model (Highest Run)

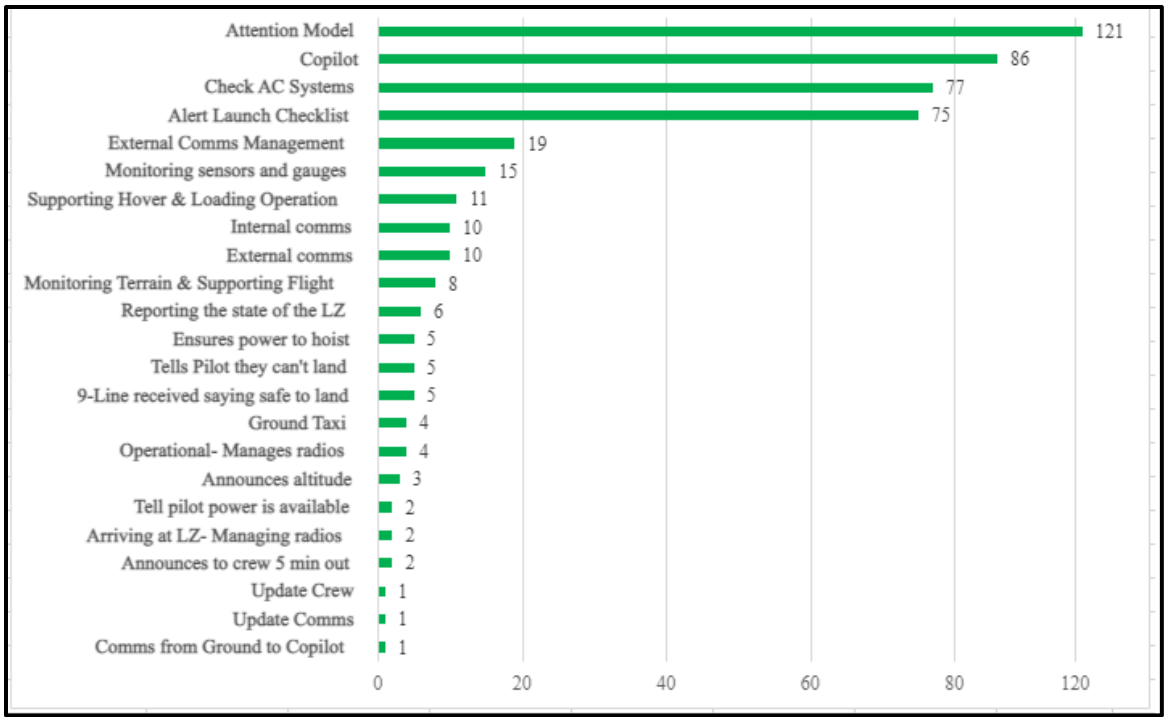

Figure 31. Co-pilot Task Over 60 SCH 1.0 Model (Highest Run)

The SCH pilot 1.0 model had two instances where cognitive workload exceeded 60 in its highest run compared to the SST pilot 1.0 model, which had 179 instances. The team determined there were not enough data points to incorporate a graph for the SCH pilot 1.0 model. The SST co-pilot 1.0 model produced fewer tasks that caused workloads over 60 (256) than the SCH co-pilot 1.0 model (473). Figures 30 and 31 describe the tasks performed in the SST Co-pilot 1.0 and SCH Co-pilot 1.0 models. The SCH Co-pilot list has 25 tasks listed, whereas the SST has 19 that exceed a workload of 60. Both figures have similar tasks, such as "Update Comms" and "Supporting Hover & Loading Operation." However, there are also some unique tasks in each figure. For example, the SCH Co-pilot list includes tasks such as "Tells Pilot they can't land" and "Check AC Systems," which are absent in the SST Co-pilot list. On the other hand, the SST Co-pilot

list includes tasks such as "Determine wind direction and velocity" and "Identify hazards," which are not present in the SCH Co-pilot list. In terms of the frequency of occurrence of tasks, the "Attention Model" is the most frequent task in both lists, with 121 occurrences in the SCH Co-pilot 1.0 model and 62 occurrences in the SST Co-pilot 1.0 model. The SCH pilot 1.0 model had two instances where cognitive workload exceeded 60 in its highest run compared to the SST pilot 1.0 model, which had 179 instances.

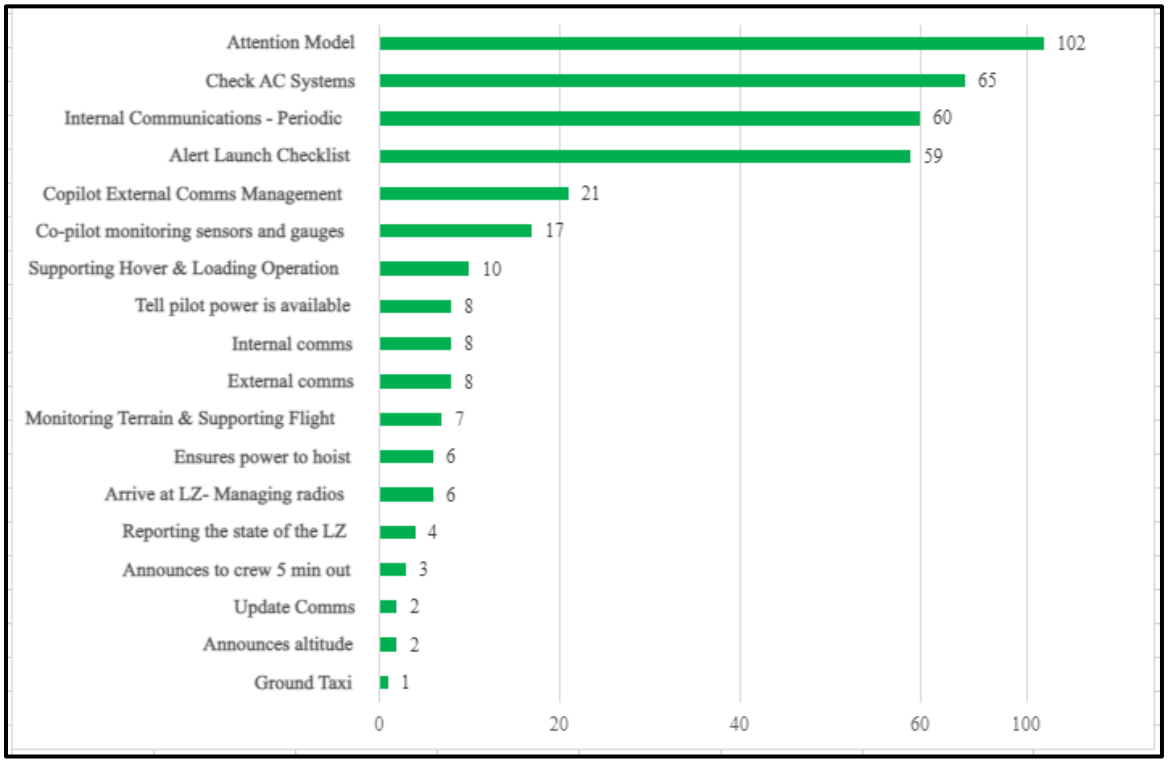

Figure 32. Co-pilot Task Over 60 SCH 2.0 Model (Highest Run)

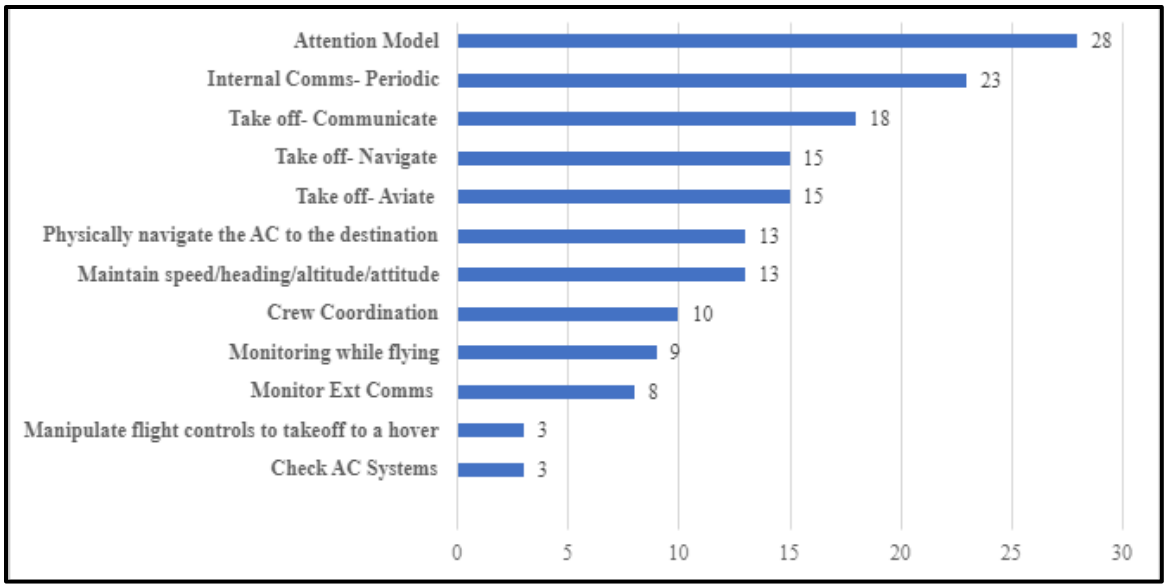

Figure 33. Pilot Task Over 60 SST 2.0 Model (Highest Run)

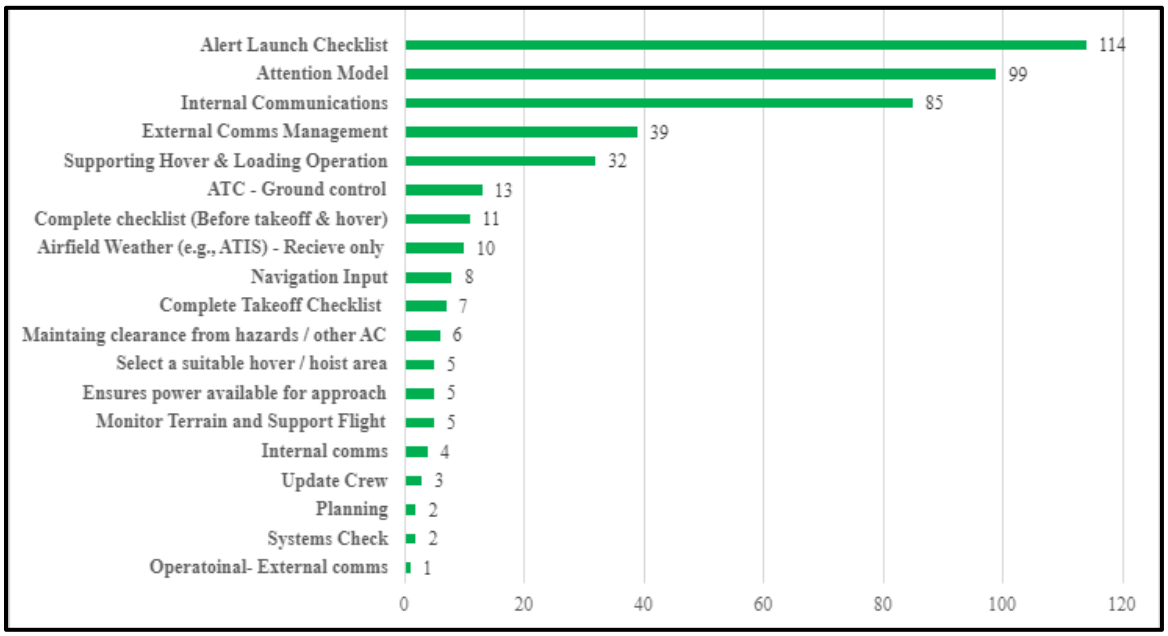

Figure 34. Co-pilot Task Over 60 SST 2.0 Model (Highest Run)

Similarly to the 1.0 pilot models, the SCH pilot 2.0 model had four instances where cognitive workload exceeded 60 in its highest run compared to the SST pilot 2.0 model, which had 48 instances. The team determined there were not enough data points to incorporate a graph for the SCH pilot 2.0 model. The co-pilot in the SCH 2.0 model

had a lower number of instances where their cognitive workload was over 60 (452) compared to the co-pilot in the SCH 2.0 model (389). Figure 32 and Figure 34 describe the tasks performed in the SCH Co-pilot 2.0 and SST Co-pilot 2.0 models. The SCH co-pilot 2.0 model has 17 different tasks identified where the workload exceeded a cognitive workload of 60, whereas the SST co-pilot 2.0 model lists 26 tasks. Both figures have some similar tasks, such as "Supporting Hover & Loading Operation" and "Internal comms." However, there are also some unique tasks in each list. For example, the SCH co-pilot 2.0 model includes tasks such as "Ground Taxi" and "Tell pilot power is available," which are absent in the SST co-pilot 2.0 model. On the other hand, the SST co-pilot 2.0 model includes tasks such as "Alert Launch Checklist (CoPilot)" and "Outside the Window," which are not present in the SCH co-pilot 2.0 model. In terms of the frequency of occurrence of tasks, the "Attention Model" is the most frequent task in the SCH co-pilot 2.0 model, with 102 occurrences, while in the SST co-pilot 2.0 model, the "Alert Launch Checklist" is the most frequent task with 114 occurrences.

F. SUMMARY

We have accomplished the task of providing the relevant data to answer the research questions by executing the methodologies discussed in Chapter III. Our analysis of the outcomes derived from four distinct models allowed us to distinguish the differences between the SCH and SST model performances and to appraise their effects on pilots' cognitive workload as represented in the models. The data shows that the inclusion of subtasks leads to a remarkable surge in instances where the cognitive workload of pilots exceeds 60. At the same time, there is a substantial decrease of 43% in instances where the workload for the co-pilot exceeds 60. It was further revealed that automating subtasks did not induce any significant change in the average cognitive workload of pilots. However, it did reduce the average cognitive workload of co-pilots by almost 9%.

Moreover, subtasks allowed for a more comprehensive depiction of the workload experienced by the pilot and the co-pilot, in contrast to the less detailed representation of workload found in lower granularity models. In addition, the VACP scores of pilots in the

SST baseline model were significantly higher than the SCH baseline model. Lastly, the VACP scores of co-pilots in the SST baseline model were relatively like those of the SCH baseline model.

Overall, the cognitive workload of pilots and co-pilots in the SCH 1.0/2.0 and the SST 1.0/2.0 models exhibit considerable similarities, with only slight variations that depend on the specific scenario.

THIS PAGE INTENTIONALLY LEFT BLANK

V. DISCUSSION

A. OVERVIEW

The primary purpose of our study was to determine the impact of model granularity and automation on the cognitive workload of pilots during a MEDEVAC mission. In this chapter, we will interpret and discuss the results presented in Chapter IV to answer our research questions. Our data elicitation and IMPRINT modeling provided the data needed to utilize statistical methods to make well-informed inferences regarding the relationship between model granularity measuring cognitive workload model granularity and the introduction of automation. Our results and observations are intended to supplement previous research teams' efforts to inform decisions regarding the development of automated systems in support of the Army's modernization initiative. Also, in this chapter, we will present and discuss the limitations and constraints we encountered that restricted our overall effort.

B. RESEARCH QUESTION ONE: DOES COGNITIVE WORKLOAD DIFFER IF MORE GRANULARITY (SUB-TASKS) IS INCORPORATED INTO A COGNITIVE MODEL?

Our research team found that the cognitive workload between a low and high granularity model differed in certain aspects. The most noticeable difference between the two 1.0 models was the pilot's workload spikes over 60. Overall, out of the ten model runs, the SCH 1.0 pilot only experienced four total workload spikes over 60 with a median value of zero. Whereas the SST 1.0 pilot experienced a total of 1,447 workload spikes over 60 with a median value of 162. The co-pilot in SCH 1.0 experienced 857 total workload spikes over 60 with a median score of 84, while the SST 1.0 co-pilot experienced 529 spikes over 60 with a median value of 45. The overall workload average between the operators in the high and low granular models was relatively similar, as shown in Figure 35. The SST 1.0 pilot has the highest average workload at 39.39, while the SCH 1.0 pilot has an average workload of 31.66. The SST 1.0 co-pilot has an average workload of 30.08, and the SCH 1.0 co-pilot has an average workload of 32.26. The

largest difference in workload is between the SST 1.0 and SCH 1.0 pilots, which is a difference of 7.73.

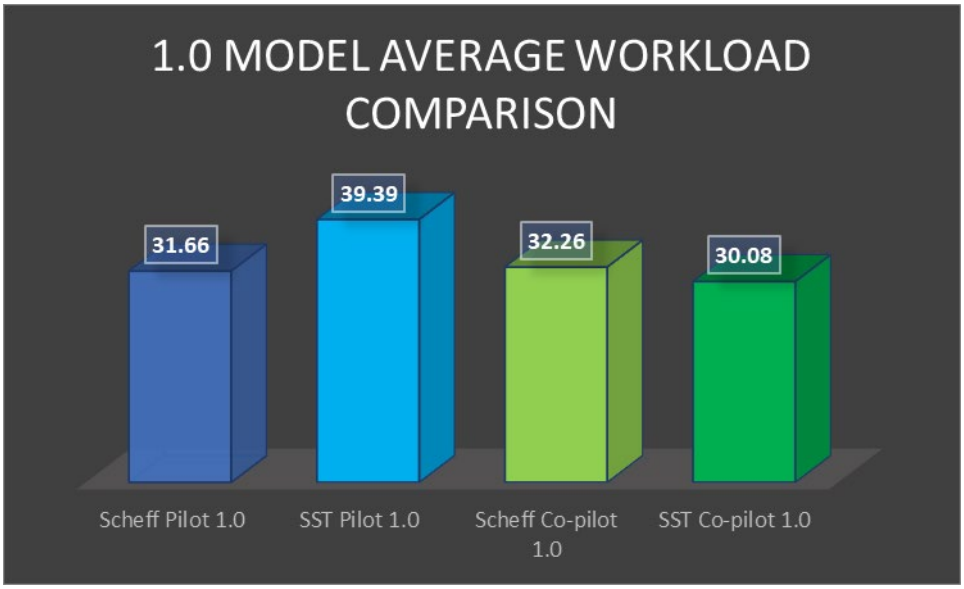

Figure 35. Graphic Comparison of Average Workload Scores of SCH 1.0 and SST 1.0, No Automation (10 Runs)

These results suggest that there is a difference in workload values and workload spikes between models that contain high and low task granularity. Even though our team added a similar number of tasks to the baseline model for each operator, that being 26 pilot tasks and 24 co-pilot tasks, the workload differences between the operators are not similar. The pilot workload differences far outpaced the co-pilot differences, and the average co-pilot workload in SST 1.0 even decreased.

The research team believes that a possible reason for the significant differences between the SST 1.0 and SCH 1.0 pilots is specifically related to task granularity and how IMPRINT calculates workload. Our research team believes that the workload difference between the two models is based on the additional task granularity that our team added and the way these tasks were added. Throughout most of the Scheff model, the pilot often only performs one single high-level task throughout the entire phase or multiple sequential tasks. The way that IMPRINT calculates workload is through the

VACP scores provided for each task. If the pilot only performs one task throughout the phase, then the VACP scores provided for that task will be the only scores contributing to the operator's workload.

The action of performing tasks simultaneously will cause IMPRINT to add the workload values together and provide an overall higher workload value for the pilot across the phase in the SST 1.0 model when compared to the single task in the SCH 1.0 model. When our research team conducted our task analysis and knowledge elicitation with pilots for our research, they often expanded the original high-level task in the Scheff model into multiple tasks. These tasks occur simultaneously multiple times in the SST 1.0 model for the pilot. When tasks occur simultaneously in IMPRINT, the model adds the workload scores together to provide an overall workload score for that operator at that given time. An example of this is found in the Perform Hover Flight task. Figure 36 shows the SCH 1.0 pilot task for performing hover flight: "Hover." This is the only task that the pilot performs during this phase. In contrast, Figure 37 shows the SST 1.0 pilot tasks that our pilots provided to add granularity to the "Hover" task. What was originally one single task became seven tasks. One of those tasks, "Manipulate flight controls to take off to a hover," serves in a loop throughout the entire phase. This means that once completed, this task will loop back to the start and run again until the phase is completely over. Having the task loop in this way will cause the pilot to perform a minimum of two tasks simultaneously throughout the phase. Conducting tasks simultaneously will lead to more workload spikes over 60, which explains why the SST 1.0 pilot has hundreds of workload spikes over 60, and the SCH 1.0 pilot only has four.

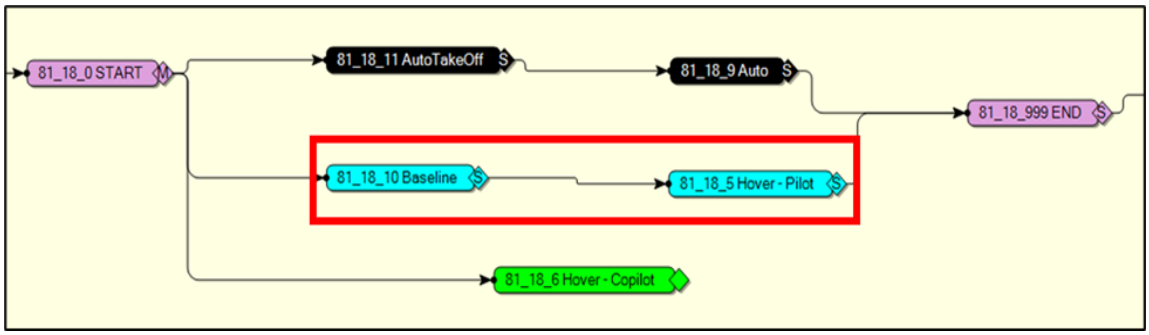

Figure 36. Scheff 1.0 Model Perform Hovering Flight Task

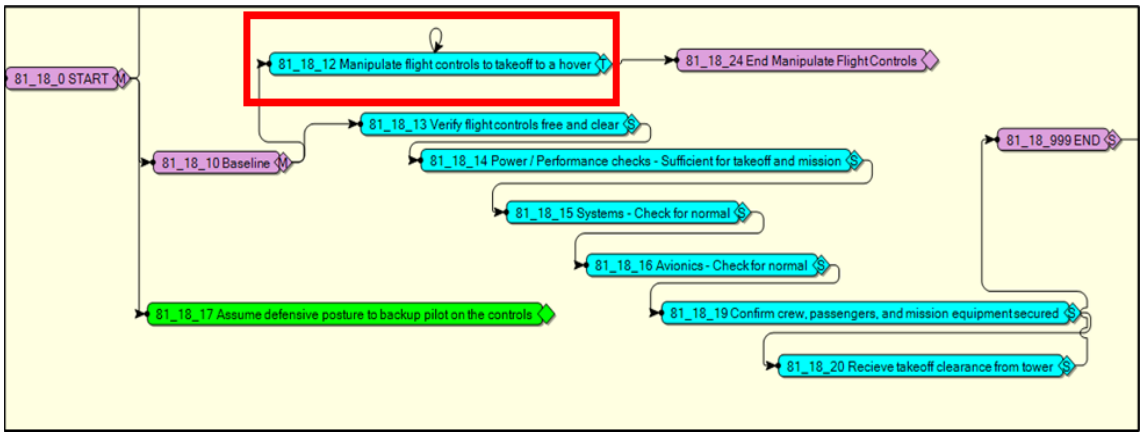

Figure 37. Seeing Sounds 1.0 Perform Hovering Flight Task with new Terminating Tasks

There are numerous tasks throughout the models where this same situation occurs, and our team believes it is the underlying reason for the difference in pilot workloads between the models. In terms of the co-pilot tasks that our team added to SST 1.0, those tasks were much more sequential than simultaneous based on the knowledge elicited from the pilots. The fact that the tasks are primarily sequential in the model means that the workload occurs sequentially as well. The co-pilot workload occurring in a sequential vice simultaneous manner could be the reason for the minor differences in co-pilot workload between the models. Finally, as shown in Figure 38, on average, the pilots rated the SST 1.0 pilot's average VACP score higher than the SCH 1.0 pilot and the SCH 1.0 co-pilot's average VACP score higher than the SST 1.0 co-pilot. These values, specifically for the co-pilots, may also explain the difference in average workload scores between the two models.

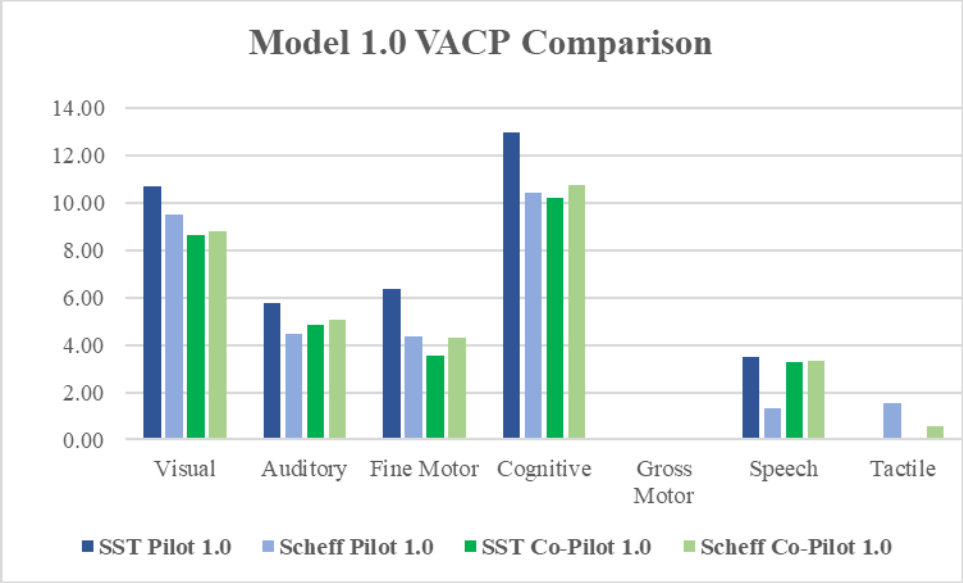

Figure 38. Graphic Comparison of Average VACP Scores of SCH 1.0 and SST 1.0, No Automation (10 Runs)

C. RESEARCH QUESTION TWO: WHAT IS THE IMPACT ON COGNITIVE WORKLOAD IF AUTOMATION IS INTRODUCED INTO A COGNITIVE MODEL?

Based on our analysis, automation has a positive impact on cognitive workload when introduced into a model. We analyzed the overall average workload scores, workload spikes over 60, and VACP results to interpret the impact of incorporating automation into a cognitive model. Our first metric, the overall average workload, shown in Table 13, provided us with a baseline workload value without automation (red box) for comparison against the average workload after introducing automation (blue box). Of the four categories in the table, the addition of automation had a marginal impact on the SCH pilot and co-pilot and SST co-pilot, with the largest reduction in workload occurring in the SST pilot. The cognitive workload of the SST pilot dropped from 39.39 to 30.45, a decrease of 23%. The rise in the overall average workload of the pilot in SST 1.0 can be attributed to the additional tasks added to the SCH 1.0 model. Because the SCH 1.0 and 2.0 are low-granularity models consisting of mostly high-level tasks that occur sequentially, the models were not adequate for simulating the effects of multitasking on workload. In the SST 1.0 model, not only were the tasks updated, but many of them were

modeled to occur simultaneously to simulate pilot multitasking, filling the gap in the SCH 1.0 model. A meaningful drop in workload was recorded after incorporating multitasking into the model and introducing automation. The decline in the overall average workload for the SST 2.0 pilot is attributed to multiple tasks being automated simultaneously, resulting in a sharp decrease in percentage. The reduced overall workload average reflects the positive impact of automation on alleviating mental workload. This is the result we anticipated based on the research findings of Parasuraman et al. (1992), Abreu et al. (2022), and Pop and Michelson (2018). Each of these studies documented the positive effect of automation on either workload or safety, and their results informed our expectations.

Table 13. Average Cognitive Workload, without and with automation

Overall Average Workload					
	SCH		SST		Overall Combined Average
	Pilot	Co-pilot	Pilot	Co-pilot	
1.0	31.66	32.26	39.39	30.08	33.35
2.0	30.13	32.41	30.45	29.62	30.66
% change	5.0	+.5	23	1.5	8

Another interesting data point in Table 13 is that the non-automated baseline (red) and automated values (blue) are generally consistent except for the average workload level of the pilot in SST 1.0. Although we anticipated an increase in the pilot's workload, we did not anticipate the values to be sharply higher when compared to the other results. Based on the previous research of Abreu et al. (2022) and input from our SMEs, we anticipated the co-pilot would have a higher workload average because of the number of tasks they are responsible for during flight. The data in the SCH 2.0 model (Table 14) confirmed Abreu et al. findings and our SME input. In SCH 2.0, the average co-pilot workload spikes over 60 occurred 71 times per flight versus nearly zero for the pilot. The

higher co-pilot number can be interpreted as the co-pilot being task saturated during flight while the pilot is not. Our results were counter to the previous research (Abreu et al. 2022). The pilot in our model had the highest overall average workload score and also had a significant increase in workload spikes. We attribute the higher-than-expected pilot values to the high-granularity IMPRINT model created by SST. Due to differences in the IMPRINT model design, the pilot tasks in the SCH model are generally sequential versus the SST model, where tasks frequently occur concurrently. Although the number of pilot and co-pilot tasks added to SST 1.0 were nearly the same, the nature of the tasks was different, which resulted in a workload spike for the pilot and minimal impact for the co-pilot. The pilot's workload increased because the updated tasks in the high-granularity model are occurring at the same time requiring more cognitive demand. The tasks added for the co-pilot in the high-granularity model were assigned as sequential tasks and did not result in a workload increase. Because of the increased demand for cognitive resources, the pilot's average workload and spikes increased. The increased cognitive demand for the pilot is also reflected in Table 14 and Table 15, where the average number of spikes increased from 0.5 in SCH 2.0 (red box) to 39.5 in SST 2.0 (blue box), with the total number of occurrences rising from 5 to 395, respectively.

Table 14. Average Number of Workload Spikes SCH 2.0 (10 runs)

	SCH Pilot 2.0			SCH Co-Pilot 2.0		
	# over 60	**Highest (MAX) Value**	**Overall average workload**	**# over 60**	**Highest (MAX) Value**	**Overall average workload**
Average	0.5	56.06	30.13	71	92.14	32.41
Total Count	5			705		

Table 15. Average Number of Workload Spikes SST 2.0 (10 runs)

	SST Pilot 2.0			SST Co-Pilot 2.0		
	# over 60	**Highest (MAX) Value**	**Overall average workload**	**# over 60**	**Highest (MAX) Value**	**Overall average workload**
Average	39.5	73.64	30.45	58	77.5	29.62
Total Count	395			580		

Again, this increase was counter to our expectations, but the result does align with Moray's theory of the relationship between resource capacity, mental workload, and the performance of complex tasks (1979). Due to the nature of the simulation, the research team could not determine if the pilot's ability to perform the tasks declined. Still, the number of workload spikes over 60 suggests the likelihood of cognitive overload, which typically leads to reduced performance. Another point of discussion regarding the pilot's average workload is the impact of automation on the overall average workload score. The overall average workload of the pilot decreased from 39.39 in SST 1.0 to 30.45 in SST 2.0, nearly a 23% reduction, as reflected in Table 13. Once again, we observe the largest decrease in the median average workload score for the pilot between SST 1.0 and 2.0 compared to the other values. We attribute this to the concurrent nature of certain tasks, which is based on the knowledge elicitation from our participants. Because our model frequently has multiple coinciding tasks, the effect on workload is more significant in the SST model when numerous tasks are automated simultaneously. Our population, consisting of all pilots, could have inadvertently focused on automating their tasks and neglected to elect additional co-pilot tasks for automation which could have unintentionally biased our results.

The last theoretical concept our analysis should consider in our results is the negative impact of mistrust on VACP values and automation. Our results were affected by the knowledge elicitation participants' lack of confidence in automated systems. In some instances, the participants reassessed the VACP value as higher than they initially reported after recommending the task for automation. The increase in the VACP score

could signify a lack of familiarity with automated systems or trust concerns with automated tasks, as discovered while conducting our literature review. The research by Parasurman et al. (1992) and Lee and See (2004) explored how a lack of trust in automation would still require a level of supervision by the user values for tasks they selected for automation, thus resulting in the allocation of additional mental resources to monitor the job. Further exploration of the role of mistrust which is beyond the scope of our research, could provide insight into the impact of automation on cognitive workload.

D. RESEARCH QUESTION THREE: IS THERE A DIFFERENCE IN COGNITIVE WORKLOAD AS A FUNCTION OF AUTOMATION AND GRANULARITY LEVELS?

As automation increases for the pilot, the amount of cognitive workload stays the same for the SCH model but decreases significantly for the SST pilot. As automation increases for the co-pilot, cognitive workload decreases for the SCH model but increases in the SST model. As the level of granularity increases, the amount of cognitive workload and spikes increases for both the pilot and the co-pilot. As detailed in Chapter Three, the SCH cognitive workload model, pilots' baseline workload spike occurrences are very low, with only four occurrences in SCH 1.0 and five in SCH 2.0. On the other hand, the baseline workload spike occurrences for co-pilots are much higher, with 851 occurrences in SCH 1.0 and 705 occurrences in SCH 2.0. The median workload spike occurrences for pilots are 0 in both SCH 1.0 and SCH 2.0, while for co-pilots, the median workload spike occurrences are 84 in SCH 1.0 and 64 in SCH 2.0. On the other hand, in the SST cognitive workload model, the baseline workload spike occurrences for pilots are much higher compared to the baseline workload spike occurrences for co-pilots. For pilots, there are 1447 occurrences in SST 1.0 and 395 occurrences in SST 2.0. For co-pilots, there are 529 occurrences in SST 1.0 and 580 occurrences in SST 2.0. The median workload spike occurrences for pilots are also higher than co-pilots in SST 1.0 and SST 2.0. Our team's findings align with Parasuraman et al.'s findings. Parasuraman et al. (1992) discovered that there is a belief among pilots that automation in the cockpit has increased rather than lowered their cognitive workload.

When examining the effects of automation on the pilots in the SCH and SST models, automation significantly impacted the SST model more than the SCH model. With the introduction of automation, the pilot in the SCH model increased cognitive workload spikes by one compared to a 1,052 decrease in cognitive workload spikes in the SST model. The team then analyzed the effects of automation on the co-pilot in the SCH and SST models. With automation, the co-pilot's cognitive workload spikes decreased by 146 in the SCH model, while the co-pilot's cognitive workload increased by 51 in the SST model. The data suggest that as we add automation into the SCH model, cognitive workload increases for the pilot while it decreases for the co-pilot. Conversely, in the SST model, the data suggest that automation decreases the pilot's cognitive workload while increasing the co-pilot's cognitive workload.

The workload spike occurrences differ between the SCH and SST cognitive workload models and between pilots and co-pilots within each model. As the model granularity level increased, the pilot total and median workload spike occurrences increased exponentially compared to the less granular models. On the other hand, for the co-pilots, as the granularity level increased, the co-pilot total and median workload spike occurrences decreased compared to the less granular models. The data for the SCH and SST models indicate that as models become granular, interviewees determined that the cognitive workload and VACP values of those tasks are higher than tasks in the low granular models. Additionally, pilots we interviewed determined that those higher granular tasks would be performed by the pilot rather than the co-pilot. Our findings are in line with Parasuraman et al.'s, which showed that there is a belief among pilots that automation in the cockpit can increase rather than decrease cognitive workload. The team has concluded that automation may decrease the cognitive workload of the pilot in a substantive manner while it may increase the cognitive workload of the co-pilot depending on how the tasks are allocated, particularly in a more granular IMPRINT model.

E. RESEARCH QUESTION FOUR: WHERE IN A MEDEVAC MISSION, AND TO WHAT EXTENT, CAN PILOT AND CO-PILOT COGNITIVE WORKLOAD BE REDUCED BY INTRODUCING AUTOMATION TO THE SYSTEM AT A SUBTASK LEVEL?

In the SST models, the overall pilot workload was significantly reduced by the introduction of automation into the model. However, the co-pilot was impacted in a different way. Overall, the co-pilot workload increased with the introduction of workload because of a significant difference in one category, "Supporting Hover & Loading Operations."

The pilot was impacted by the following tasks that the SMEs recommended automating:

- Maintain desired airspeed and heading
- Continues to fly
- Physically navigate the aircraft to destination
- Avoid traffic obstacles
- Manipulate flight controls and take off to a hover
- Aviate
- Navigate

In automating these seven tasks in the SST model, there was an 82% decrease in the number of times these tasks spiked over 60, from 333 to 59 tasks. This reduction in tasks associates with a reduction in cognitive workload. In two categories, "Continues to fly" and "Avoid traffic obstacles," automation drove these tasks below 60. These pilot tasks appear in the "Location Update from Ground" and "Low-level (flying fast)" phases, respectively.

Regarding the co-pilot, the automation impact was not as forthright. The co-pilot was impacted by the following tasks that the SMEs recommended automating:

- Identify flight path corrections
- Identify hazards
- Identify navigation checkpoints
- Monitor terrain and support flight
- Support hover & loading operation

The number of times the operator performed these tasks where their workload spiked over 60 decreased by 63% in the first four tasks, from 16 tasks to 6 tasks. "Identify flight path corrections" and "Identify navigation and checkpoints" both fully alleviated cognitive workload spikes above the threshold of 60. These tasks both occur in the "Low level (flying fast)" phase of flight. However, when it comes to the "Support hover & loading operation" task, the number of times the task appears over 60 jumps from 6 to 32 times when automation is introduced. This is an increase of over 400%.

This workload spike in the SST 2.0 model could have occurred for at least two reasons. First, the VACP values increased between the SST 1.0 and the SST 2.0 models. In the SST 1.0 model, the VACP values used came from the median value among the five data sets the research team received from the knowledge elicitation. In the SST 2.0 model, only one pilot recommended that "Support hover & loading operation" be automated. When incorporated into the model, this individual recommended VACP scores higher than the median value. As a result, the visual VACP score increased from 3.35 to 5.00, the fine motor value increased from 2.2 to 5.00, and the tactile value increased from 1.0 to 4.0. Our research team discussed this observation with the pilot. He explained to the team that he assumed that if this task were automated, he would still need to maintain vigilance visually and assume a tactile response would be added to the system to confirm that the automation had completely taken over the task.

The pilot's response parallels research the team found in our literature about trust in automation. In order for the pilot to trust the automated task, the operators need reassurance that the aircraft is operating effectively without their manual control. Visual,

audio, or tactile feedback can improve the operator's situational awareness to ensure the automation is working as designed (Endsley and Kiris 1995). Our literature also confirms that this introduction of automation introduces the paradox that automation may unintentionally increase cognitive workload due to the need to monitor the system constantly, as Parasuraman et al. (1992) found. This is why the tasks recommended for automation need to be carefully considered to not impact automation adversely.

Second, the SST 2.0 spike increase may be due to how the model was created. The data we used to analyze these overload tasks came from a model run that had the most workload spikes. The standard deviation of task time may have impacted "Support hover & loading operation" because it runs as a loop in the model. This task runs while the pilot is completing three tasks in sequence. Each of these tasks runs for eight minutes, with a standard deviation of four minutes. That means that in any run, there is a variation between a minimum of 12 minutes and a maximum of 36 minutes in the loop. When the pilot completes the third task, "Hover during litter raise," the looped task stops. The longer it takes the pilot to complete their tasks, the longer it takes the co-pilot to finish this task. The longer it takes the co-pilot to complete the task means that it is impacted more by the attention model, which drives up the cognitive workload. This result may not be as extreme in another run of the same IMPRINT model because this tool uses stochastic discrete-event simulation, which generates random variates in each run. This means that IMPRINT can simulate multiple mission possibilities by capturing the randomness and probability of choices made by the operators.

In the SCH models, the overall average pilot cognitive workload decreased by 5%, changing from a 31.66 to a 30.13 workload value with the introduction of automation. There were 0 spikes over 60 for this operator in both models, so no tasks were identified in overload.

For the co-pilot, the overall average cognitive workload difference was negligible, increasing cognitive workload by 0.5% (from 32.26 to 32.41 workload value) with the introduction of automation. The number of tasks that appear in overload was reduced from 473 tasks to 389, or 18% overall, completely alleviating five tasks between the SCH 1.0 and the SCH 2.0 model:

- Communication from Ground to Co-pilot
- Update Crew
- Operational- Manages Radios
- 9-Line received saying safe to land
- Tells Pilot they can't land

The tasks "Communication from Ground to Co-pilot," "9-Line received saying safe to land," and "Tells pilot they can't land" are all impacted by automation in SCH 2.0. When Mr. Scheff incorporated automation into his model, it specifically automated pilot tasks, not co-pilot tasks. However, when the pilot tasks are automated, it bypasses these identified co-pilot tasks. Consequently, this may have helped decrease the workload by excluding tasks.

F. LIMITATIONS

A significant impediment experienced by the team throughout the study was a need for more organic technical expertise in avionics. Because none of our team members are qualified pilots and have no academic background in aeronautics, our initial knowledge of the tasks performed by pilots and their flight sequences was limited. Although the team lacked operator experience, we have conducted a considerable number of training exercises and combat missions in support of military operations that included helicopters. We mitigated our knowledge gap by reviewing the United States Army Aviation Center of Excellence UH-60 Series Aircrew Training Manual (DA 2022) to familiarize ourselves with the tasks conducted by operators. This manual provides the master task list for UH-60 pilots, which also served as the basis for the tasks in Mr. Scheff's model. Our team reviewed this manual to gain knowledge about the tasks pilots are expected to execute while performing their missions. We enhanced our understanding of the mission tasks during conversations with two subject matter experts who helped shape our knowledge elicitation script and tasks that we incorporated into our model.

Our research team found that using a pre-existing model was both a benefit and a limitation. Thanks to Mr. Scheff, the research benefitted from having a fully functional model as a starting point rather than creating a model from the ground up. Mr. Scheff's model served as an excellent starting point for the beginning of our task analysis as it was already structured around a MEDEVAC flight. Without utilizing the Scheff model as a foundation, the scope of our research would have been limited due to the amount of time required to build the simulation. Conversely, the team encountered issues adapting the model to our purpose because Mr. Scheff built his model for his specific use. Because our goal did not align with his, we had to account for how the model was created to support its original intent as we progressed through knowledge elicitation. The team successfully adapted the model with the help of Mr. Sargent, who diligently guided us through the intricacies of Mr. Scheff's advanced model as we created our revised models. A significant issue our team encountered during modeling that required Mr. Sargent's expertise was accounting for the attention model and communication tasks that ran passively in the model's background. These background tasks ran throughout the model, and we had to exercise care on which tasks we included in the model to mitigate possibly duplicating workload. Finally, Mr. Scheff utilized a large amount of code throughout his model to trigger different paths or actions to occur. It is likely that without more IMPRINT training or the readily available help of Mr. Sargent, our team would have failed to use Mr. Scheff's model correctly. Future research teams should weigh the pros and cons of using a previously validated model against the time and effort required to build their own.

Another limitation faced by our team was limited participation from subject matter experts. Our team initially identified ten pilots who indicated they would participate in the knowledge elicitation phase of the project. However, only five pilots responded to requests to participate in the study. This small sample size negatively impacts our ability to attest to the accuracy and relevance of our data and results. Unfortunately, the small pool of only five participants limits our ability to interpret results and draw conclusions. Additionally, all five pilots that responded are currently pilots or had recently served in the position. None of them had recent co-pilot experience.

This lack of participation by co-pilots serving in that capacity may indicate a bias towards pilot tasks and workload. Future research teams should strive to include actively serving co-pilots to ensure their perspective is accounted for. The lack of warrant officer participation in the project was another limitation. The five pilots who responded were all commissioned officers, leaving our data without input from warrant officers. Warrant officers are a crucial population in the MEDVAC world and are often the most experienced operators in the aviation branch. Future research teams should also strive to include warrant officers in their sample size to account for their depth and breadth of knowledge and experience.

Finally, the most significant limitation faced by our team during the study was simply time. The initial model provided by Mr. Scheff contained multiple paths that the flight could take as well as different scenarios, such as receiving enemy fire or being re-directed mid-flight to a new pick-up point. Our team did not have the time to explore these different model paths and scenarios and were limited to only one path. Future research teams may benefit from exploring different paths in the Scheff model to continue to explore the difference between high and low-granularity models.

G. SUMMARY

One of our goals for this capstone was to identify the human capabilities and limitations that can be augmented by the emergent technologies being developed in support of the FVL program to achieve total systems integration. To achieve our goals, we needed to determine if there was a relationship between the introduction of automation and workload reduction and provide an enhanced IMPRINT model that our stakeholders could utilize to conduct future research. Following the analysis of our results, the data indicate that the aircrew is under a considerable cognitive load when multitasking, especially when dedicating substantial mental resources to simultaneously conducting visual and cognitive tasks. Our results also imply that the pilots' workload is affected after introducing an automated agent. The results also revealed the differences between high-granularity and low-granularity IMPRINT models for accurately simulating workload. This result signifies that a high-granularity IMPRINT is an enhancement

compared to a low-granularity model. In terms of our research objectives, we have achieved our goals, and although we encountered several limitations and constraints, we maximized the tools available to deliver to our primary stakeholder valuable cognitive workload research and an updated IMPRINT model.

THIS PAGE INTENTIONALLY LEFT BLANK

VI. CONCLUSION AND RECOMMENDATIONS

A. OVERVIEW

The primary purpose of this capstone project was to introduce automation into a previously validated IMPRINT model and determine the effect of automation on the cognitive workload of a UH-60 MEDEVAC pilot and co-pilot during a MEDEVAC flight. The research team also determined if there was a difference in workload values across a simulated MEDEVAC flight from the perspective of a high-level model and another model with the same high-level tasks but with more detailed sub-tasks included. The research team also sought to determine if there was a significant difference in workload values between these two models, both with and without the introduction of automation. This study has examined four research questions:

- Does cognitive workload differ if more granularity (subtasks) is incorporated into a cognitive model?
- What is the impact on cognitive workload if automation is introduced into a cognitive model?
- Is there a difference in cognitive workload as a function of automation and granularity levels?
- Where in a MEDEVAC mission, and to what extent, can pilot and co-pilot cognitive workload be reduced by introducing automation to the system at a subtask level?

Of the 39 tasks involved in the cognitive workload spikes over 60, 12 were identified by the subject matter experts as candidates for automation. We discovered that overall, the cognitive workload of the operators in both models show considerable similarities, with only slight variations that depend on the specific scenario. Of note, we did notice that the pilot workload average and the number of spikes above 60 in the SST models were significantly higher when compared to the SCH models.

B. CONCLUSIONS

This research demonstrated that there does not appear to be a significant difference when it comes to the average workload score both overall and in the VACP channels between a high and low granularity model. It did, however, show a substantial difference in workload spikes of over 60, at least for the pilot. Additionally, it showed that automation appeared to have a minimal impact on the overall workload for the operators while significantly reducing the pilot's workload spikes over 60.

The research further demonstrated that IMPRINT is a valuable tool for building scenarios to model the cognitive workload of operators. IMPRINT is also an excellent tool to either insert an automated agent into the model to achieve automation in the scenario or by reducing the current operator's VACP scores to replicate their reduced cognitive workload for the task if it was automated. The research provides the HSA-DM team with an option of using either a high- or low-granularity model of a simulated MEDEVAC flight and an additional option of continuing to add granularity to the high-granularity model. The research also provided the HSA-DM team with the list of tasks that occur during the workload spikes over 60 experienced by each operator as well as the frequency with which those tasks align with workload spikes over 60. This task list was then cross-referenced against the tasks the pilots recommended for automation during the knowledge elicitation. Where these two lists intersect are, in our team's estimates, the prime candidates at the task level to consider when determining how automation could be incorporated in an effort to reduce operator cognitive workload during missions.

The process of conducting knowledge elicitation and task analysis to create a higher granularity model was instrumental in identifying those tasks from the original model that could be subdivided into multiple tasks as well as identifying new tasks that needed to be added by themselves. The discussions with the pilots helped our research team to adjust our flight scenario for the model as well as how we explained the examples of the VACP channels. Talking through our scenario and explaining the intent of the research with the pilots was critical to gaining accurate information about the tasks performed, their frequency, duration, and each channel's workload value.

Internal and external communications heavily contribute to operator workload spiking over 60. This was true in all four models. In three of the four models, the cognitive workload was relatively evenly distributed amongst the pilot and co-pilot. In SST 2.0, SCH 1.0, and SCH 2.0, the workload difference between the operators was relatively small. Only SST 1.0 produced a rather large workload difference: the pilot's average workload across the ten runs was almost 10 points higher than the co-pilot's. These workload values were elicited from study participants based on the specific flight scenario conditions provided to them. Should HSA-DM wish to see this model played out under different conditions and assumptions, it would only require an update to the flight scenario and pilots to provide new VACP scores based on the new scenario.

Adding automation to the models revealed that, in some cases, the pilots provided higher VACP scores in some channels for the automated tasks compared to the non-automated task. This implies that some pilots view monitoring an automated agent performing certain tasks as requiring a higher cognitive workload than if they were to do the task themselves. Surprisingly, the co-pilot's average workload in the Scheff models increased, albeit slightly, when adding automation. Table 16 shows that the co-pilot's average workload in SCH 1.0 was 32.26, increasing to 32.41 in SCH 2.0. The largest impact that automation had in our models was on the pilot's workload score in SST 1.0 and 2.0. The pilot's average workload score in SST 1.0 was 39.39, which, after introducing automation in SST 2.0, was reduced to 30.45. This is likely because the pilots in the knowledge elicitation phase of the research identified several tasks as automation candidates that contained high VACP scores and were longer than most in duration.

Table 16. Comparison of Average Workload in SCH Cognitive Workload Model, Baseline versus Automated Mode

	Overall Average Workload			
	SCH		SST	
	Pilot	Co-pilot	Pilot	Co-pilot
1.0	31.66	32.26	39.39	30.08
2.0	30.13	32.41	30.45	29.62

This project revealed some constraints with the IMPRINT modeling software. Specifically, IMPRINT does not currently have a feature that allows a user to execute multiple runs of their scenario and provide that user with the results of those runs averaged together. Should one wish to run their model multiple times to achieve a desired sample size of data, they must conduct one model run at a time and save each spreadsheet separately. Then the user must combine the spreadsheets into one set of data, allowing them to provide an average workload value across all runs. However, this approach increases the potential for data discrepancies and increases the researcher's data processing workload. If IMPRINT were updated to include a feature allowing a user to complete dozens or hundreds of simulations with one click of a button and provide the distribution of the overall workload and VACP channel workload scores, it would increase accuracy and efficiency.

This project also revealed some of the issues future researchers could encounter when attempting to use a pre-existing IMPRINT model that was purposefully designed to answer another modeler's research question. Our team encountered a few such issues attempting to adapt the pre-existing model provided by Mr. Scheff to suit our research questions. Thankfully, with the assistance of Mr. Scheff and Mr. Sargent, the team was able to identify and mitigate the issues.

Our primary concern was accounting for tasks running in the background of the model for its entire duration, which was not discovered until after we completed our

models, ran the simulation, and began to analyze the data. These tasks had to be isolated in the model and either turned off or accepted as part of the model. Mr. Scheff had several paths in his original model controlled by code written inside IMPRINT tasks. Our team had to ensure that when selecting our own path, we did not retain any other task workload on the other paths and that turning off other paths did not break the model. While there are many benefits to beginning a project with a pre-existing IMPRINT model, our team recommends that any future researchers who find themselves in this situation thoroughly explore the model with an IMPRINT SME as we did to ensure that decisions researchers make do not result in an incorrectly working model.

C. RECOMMENDATIONS FOR HSA-DM

The tasks our team identified as possible candidates for automation were based on their occurrence in operator workload spikes over 60 and if the tasks were recommended by the pilots for automation:

- Maintain desired airspeed and heading
- Pilot continues to fly
- Physically navigate the aircraft to destination
- Avoid traffic obstacles
- Manipulate flight controls to take off to a hover
- Aviate (Autopilot)
- Navigate (Autopilot)
- Identify flight path corrections
- Identify hazards
- Identify navigation checkpoints
- Monitor terrain and support flight

- Support hover and load operations

Automating these tasks in the SST 2.0 model reduced the pilot's average cognitive workload from 39.39 to 30.45, while the co-pilot's average cognitive workload was reduced from 30.08 to 29.62. When automation was introduced to the Scheff model, the pilot's average cognitive workload was reduced from 31.66 to 30.13, while the co-pilot experienced an increase in average cognitive workload from 32.26 to 32.41. Automating these tasks in SST 2.0 and using the pre-existing automated paths in SCH 2.0 decreased workload spikes of over 60 from 207 to 88 in SST and from 84 to 64 in SCH. One of the key issues remaining would be if and how automation can be introduced to these tasks. These tasks were only identified as candidates for automation without consideration of technical feasibility. Further research should be conducted to determine if these tasks are feasible candidates for automation and, if so, how automation can be introduced. Additionally, varying levels of automation should be considered to determine if automating only a part of a task is more feasible or beneficial than automating the entire task.

D. RECOMMENDATIONS FOR FUTURE RESEARCH

While the results produced by this research team and the IMPRINT model are informative, the task frequency, duration, and workload values were only provided by five of the ten pilots who participated in knowledge elicitation. Therefore, the team's first recommendation for future research would be to expand the pool of pilots to provide a greater sample size which will provide more accurate task frequency, durations, and workload values. Additionally, we recommend that care be taken to ensure that warrant officers are included in the expanded pool of pilots interviewed and that some interviewees should be currently serving in the co-pilot role to guard against a bias towards pilot tasks and workload.

Our second recommendation would be to remove the attention model tasks and the internal and external communication tasks that currently run in the background of this model. Eliminating these tasks that consistently run in the model's background and adding them as discrete tasks where necessary based on knowledge elicitation with

subject matter experts will allow the results portion of the IMPRINT model to be run with conflict values added. Conflict values allow IMPRINT to use the multiple resource theory and will greatly increase workload where tasks are conducted simultaneously rather than simply adding the VACP values of each task together as is currently done in our model. This would allow a workload comparison between an MRT model and a model based solely on VACP values.

Our third recommendation is to conduct additional analysis in the model where the workload for the operators drops to zero. In a real-life scenario, there should be no instances where the workload ever truly drops to zero. A pilot and co-pilot are continuously doing something throughout the execution of the mission. The research team recommends that future researchers conduct additional knowledge elicitation and task analysis with subject matter experts to account for these areas in the model where the workload drops to zero. The points in time where workload drops to zero should be identified and targeted for discussions with subject matter experts to determine what tasks they are executing during these periods and then incorporate those tasks into the model for a more realistic simulation.

Our final recommendation is to create separate IMPRINT models to support both future long-range assault aircraft and future attack reconnaissance aircraft missions. Developing high-granularity models to simulate the anticipated interfaces, interactions, and tasks associated with the next-generation FVL aircraft will provide valuable insight into the effects of new technology on cognitive workload. The next generation of helicopters is anticipated to fly further, faster, and longer and will place a tremendous amount of additional workload on the pilots. By incorporating IMPRINT modeling early in the design and prototyping phase, design recommendations can be implemented before it becomes cost prohibitive. We recommend convening a team that represents the project's stakeholders and beginning the development of the IMPRINT model.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX. SCRIPT FOR MEDEVAC OPERATIONAL SCENARIO

Operational Context:

- On behalf of the research team, welcome and thank you for your time
- Introduce the team
- Ask them if they are ok with us recording the interview
- Using it for a record of the conversation
- The interview will not be shared with anyone outside of the research team

Provide the purpose of the interview:

The Holistic Situational Awareness – Decision Making (HSA-DM) office is exploring introducing additional automated tasks to the Future Vertical Lift program than what is currently available in today's aircraft to reduce the cognitive workload demand on pilots that have been seen to become increasingly excessive. Previous research has shown that pilot and co-pilot cognitive workload during the steady state phase of a MEDEVAC flight exceeds the safe threshold value of 60, and that automation of specific tasks helps to reduce the frequency and time of those spikes. Our team aims to continue this research by exploring an entire MEDEVAC flight from start to finish and assessing the impact of automated tasks on cognitive workload.

The software utilized to measure cognitive load is called IMPRINT. Discuss the IMPRINT tool. The improved performance research integration tool is a modeling software where values for cognitive loads can be input, and then simulations can be run to observe the effects of multitasking.

The objective of our research is to expand and confirm the baseline for pilot and co-pilot cognitive workload values across the duration of a MEDEVAC flight and target specific tasks to automate and incorporate them into the model. We will then run the

model again and compare the workload values between the two to assess the impact of automation on the aircrew's workload.

Introduce the products.

- Walk them through a concept map and visual aid.
- I am showing you the concept diagram that is intended to represent the stages of a MEDEVAC flight and the tasks executed during those stages.
- Think of it as a road that starts at the beginning and walks us through until the end.
- I would like to highlight the legend in the upper left corner, and the different colored boxes represent whether the pilot or copilot is the primary person assigned the task.
- The tasks are numbered in such a way as to depict seniority. For example, 1.0 – Takeoff is the parent task of 1.1 – Check AC Systems, which is the parent task of 1.1.1 Check AC Systems – Pilot.

We will ask you to assign VACP values and workload scores to each of the terminating tasks that we identify. As highlighted with the red boxes.

- Introduce Excel, discussing assigning values.
- Where and how, for the first one, we will go slow.
- The Excel document follows the concept map and will be used to record the VACP values and workload scores you provide for each task.
- These values should be based on your experience.
- In addition to these scores, we would like to ask you to provide an estimate on the frequency or number of times you perform that task, the time duration that the task lasts, and an estimated standard deviation in

time that it may take you to complete the task. (Interviewer will provide an example here).

- The last column in the Excel is your recommendation to automate tasks.
- The workload score options are provided on the paper copy of the VACP values we provided to you. We will go over that next.
- However, if you believe the workload score is in between the provided numbers, state what you think it is, and we can input that value into the model.
- VACP walk-through- one per

Limitations of the study

Our model is limited because of the team's limited time to complete the project. Establishing the foundation for the next team. Ask the interviewee if the model is missing any key information. This is based on the UH-60 Aircrew Training Manual (JAN 22). We have more freedom to change the terminating tasks or the tasks in the red box.

- Review to bio
- Ask for any updates and ask about other qualifications
- Ask for the number of months deployed
- Before we move on to the tactical scenario, do you have any questions?

Tactical scenario

Background Assumptions:

- You are piloting a UH-60M.
- The crew has trained and qualified together in CONUS before deployment.

- The crew has been in the combat theater for approximately 30 days and has just completed their relief in place/transfer of authority (RIP/TOA) with the outgoing unit.
- The crew has met all requirements necessary to deploy in theater (e.g., environmental qualifications, etc.)
- You are the PC and can utilize your personal experience in flight hours and years for your basis of estimates.
- Your co-pilot is competent – you can trust that they can and will accomplish all expected tasks and duties to standard.
- The weather conditions presented in the scenario occur when you arrive at the known landing area (Phase VI)
- AH-64 Gunship escort is ready to launch, and on comms, AC requires no maintenance and is FMC

General Scenario Initiation:

A 9-Line MEDEVAC w/ MIST request is received for 1x urgent surgical patient. The patient received gunshot and shrapnel wounds from enemy fire. The pick-up location is approximately 40 kilometers away from the aviator's base of operations and 20 kilometers away from the CSH (Combat Support Hospital).

There is a potential that the aircraft may come under fire from the ground while en route to the LZ or at the LZ. (Traditional threats as seen in Iraq or Afghanistan – RPG, Small arms, etc.)

Specific Scenario Conditions:

- Atmospheric conditions:
- Visual Flight Rules (VFR)
- Ceiling – 1000ft

- Visibility – 3 Statute Miles (SM)
- Winds – 10 - 15 knots
- Precipitation – None
- Temperature – 20 Celsius
- A rainstorm is forecasted to arrive later in the evening
- Illumination is green
- The mission will occur at night (under NVGs).
- Crew members have had 10 hours of rest.
- The crew has flown multiple times for up to six hours in the past two weeks, supporting the RIP/TOA and orientation flights.
- The flight will occur in mountainous terrain and will require multiple ridgeline crossings, flying through valleys as well as accounting for other additional hazards, such as power lines, towers, etc., as you approach the CSH. The highest point during the flight is 9,000 MSL.
- The mission will involve an escort of 2 x AH-64 Apache helicopters due to the threat of enemy fire en route and on the LZ. Airspace deconfliction is necessary throughout the flight.
- Navigation is input to include MEDEVAC pickup location information
- Assume that all theatre-wide air corridors and aviation checkpoints are established
- Weather is legal – all necessary materials or items required for clearance to take off are complete

Specific Scenario Conditions at Landing Zone:

- You approach the LZ at 200’ Above Ground Level (AGL), flying into the wind to conduct an assessment.
- It is determined that the landing zone is not suitable due to terrain and obstacles, and a hoist extraction will need to be conducted at approximately 50’.
- You have Out of Ground Effect + 15% power margin with no expected performance limitations that would impede your ability to safely hover the aircraft.

SUPPLEMENTAL 1. CONSOLIDATED DATA OF VACP SCORES AND TASKS RECOMMENDED FOR AUTOMATION FROM SUBJECT MATTER EXPERTS

This supplemental document contains the combined results from the research team's knowledge elicitation phase. During this phase, we conducted cognitive walkthroughs of a MEDEVAC flight with five subject matter experts to obtain and record the data necessary to populate the IMPRINT models. The individual and aggregate results are included in this supplemental document with the aggregate values used to populate the model and obtain our results. If you would like to view this supplemental information, please direct your inquiry to the Naval Postgraduate School library.

THIS PAGE INTENTIONALLY LEFT BLANK

SUPPLEMENTAL 2. SEEING SOUNDS IMPRINT MODELS

This supplemental data file contains the four individual IMPRINT models utilized by our team during this research project. Scheff 1.0 and 2.0 are the models based on the original model provided by Mr. Scheff as a starting point to facilitate our research. These models were relatively unchanged by the research team, with the exception being that we adjusted the duration of each MEDEVAC flight phase to more closely match the phase times found in the Seeing Sounds models. Seeing Sounds 1.0 and 2.0 are the models that contain a higher granularity and are the result of the knowledge elicitation and task analysis conducted by our team with subject matter experts. If you would like to view this supplemental information, please direct your inquiry to the Naval Postgraduate School library.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Abreu, Leonardo, John Girton, Mary Lord, Matthew Miller, and Rodolfo Zelada. 2022. "Using Modeling to Predict the Effects of Automation on Medevac Pilot Cognitive Workload." Master's thesis, Naval Postgraduate School. https://hdl.handle.net/10945/70616.

Adhikaei, Pralhad. 2015. "Errors and Accidents in the Workplace." *Sigurnost* 57 (2) (February): 127–137. https://libproxy.nps.edu/login?url=https://www-proquest-com.libproxy.nps.edu/scholarly-journals/errors-accidents-workplaces/docview/1693680491/se-2.

Andrews, Jinan M., Christina F. Rusnock, Michael E. Miller, and Douglas P. Meador. 2020. "Reshaping Airpower: Development of an Imprint Model to Analyze the Effects of Manned-Unmanned Teaming on Operator Mental Workload." In *2020 Winter Simulation Conference (WSC)* 170–181. https://doi.org/10.1109/WSC48552.2020.9383859.

Broadbent, Donald E. 1958. *Perception and Communication*. New York: Pergamon Press.

Chen, Fang, Jianlong Zhou, Yang Wang, Kun Yu, Syed Z. Arshad, Ahmad Khawaji, and Dan Conway. 2016. *Robust Multimodal Cognitive Load Measurement*. Cham: Springer International Publishing AG. https://doi.org/10.1007/978-3-319-31700-7.

Carter, Matthew W., Greg Griffith, Pete Hamill, and Jacen Lanclos. 2021. "What Tasks to Automate? An Investigation of What Tasks Make Sense to Automate for Future Aviation Platforms." Master's thesis, Naval Postgraduate School. http://hdl.handle.net/10945/68802.

Endsley, Mica R., and Esin O. Kiris. 1995. "The Out-of-the-Loop Performance Problem and Level of Control in Automation." *Human Factors and Ergonomics Society* 37 (2) (June): 381–394. https://doi.org/10.1518/001872095779064555.

Eysenck, Michael. 1994. "Books Reconsidered, Perception and Communication." *British Journal of Psychiatry* 164 (5) (May): 711–712. https://doi.org/10.1192/S0007125000034759.

Federal Aviation Administration. 2022. "Fly the Aircraft First." https://www.faa.gov/news/safety_briefing/2018/media/SE_Topic_18-07.pdf

Fox, Julia R., Byungho Park, and Annie Lang. 2007. "When Available Resources Become Negative Resources: The Effects of Cognitive Overload on Memory Sensitivity and Criterion Bias." *Communication Research* 34 (3) (June): 277–296. https://doi.org/10.1177/0093650207300429.

Galant-Gołębiewska Marta, Wojciech Zawada, and Marta Maciejewska. 2020. "Analysis of Pilot's Cognitive Overload Changes During the Flight." *Advances in Military Technology* 15 (2) (September): 329–342. https://doi.org/10.3849/aimt.01408.

Galy, Edith., and Claudine Melan. 2015. "Effects of Cognitive Appraisal and Mental Workload Factors on Performance in an Arithmetic Task." *Applied Psychophysiology and Biofeedback* 40 (4) (July): 313–325. https://link-springer-com.libproxy.nps.edu/article/10.1007/s10484-015-9302-0.

Gertler, Jeremiah. 2021. *Army Future Vertical Lift (FVL) Program*. CRS Report No. IF11367. Washington, DC: Congressional Research Service. https://crsreports.congress.gov/product/pdf/IF/IF11367/6.

Gouraud, Jonas, Amaud Delorme, and Bruno Berbian. 2017. "Autopilot, Mind Wandering, and the Out of the Loop Performance Problem." *Frontiers in Neuroscience* 11 (541) (October): 1–13. https://doi.org/10.3389/fnins.2017.00541.

Guyton, Zachory J., and Richard Pak. (2020). "The Impact of Automation Etiquette on User Performance and Trust in Non-Personified Technology." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 64 (1) (December): 1805–1809. https://doi.org/10.1177/1071181320641435.

Jankowska, Maja. 2014. "Concept Mapping: A Tool of Multiple Purposes in Research." In *SAGE Research Methods Cases Part 1:* 1–17. London: SAGE Publications Ltd. https://dx.doi.org/10.4135/978144627305013514688.

Johannsen, Gunnar. 1979. "Workload and Workload Measurement." In *Mental Workload: Its Theory and Measurement,* edited by Neville Moray, 3–11. New York: Published in coordination with NATO Scientific Affairs Plenum Press.

Kahneman, Daniel. 1973. *Attention and Effort*. Englewood Cliffs, NJ: Prentice Hall.

Lee, John D., and Katrina A. See. 2004. "Trust in Automation: Designing for Appropriate Reliance." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 46 (1) (January): 50–80. https://doi.org/10.1518/hfes.46.1.50_30392.

Maclsaac, Margaret A. 1998 "The Effect of Automation on Helicopter Crew Communication: A Low-Fidelity Investigation." Master's thesis, Embry-Riddle Aeronautical University. https://commons.erau.edu/db-theses/129.

Moray, Neville. 1979. "Models and Measures of Mental Workload." In *Mental Workload: Its Theory and Measurement*, edited by Neville Moray, 13–22. New York: Published in coordination with NATO Scientific Affairs, Plenum Press.

National Aeronautics and Space Administration. 2020. "NASA-STD-3001 Technical Brief" https://www.nasa.gov/sites/default/files/atoms/files/cognitive_workload_technical_brief_ochmo.pdf.

National Transportation Safety Board. 1994. . Report No. AAR-94/07. https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR9407.pdf .

Parasuraman, Raja, Toufik Bahri, John E. Deaton, Jeffrey G. Morrison, and Michael Barnes. 1992. *Theory and Design of Adaptive Automation in Aviation Systems*. Report No. NAWCADWAR-92033-60. Washington, DC: Naval Air Warfare Center. https://apps.dtic.mil/sti/citations/ADA254595.

Parasuraman, Raja, Robert Molloy, and Indramani Singh.1993. "Performance Consequences of Automation-Induced "Complacency." *The International Journal of Aviation Psychology* 3 (1) (February): 1–23. https://www-tandfonline-com.libproxy.nps.edu/doi/epdf/10.1207/s15327108ijap0301_1?needAccess=true&role=button.

Parasuraman, Raja, Thomas B. Sheridan, and Christopher D. Wickens. 2000. "A Model for Types and Levels of Human Interaction with Automation." *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans* 30 (3) (May): 286–297. https://doi.org/10.1109/3468.844354.

Pond, Travis, Brandon Webster, John Machuca, John Colombi, Michael Miller, and Randall Gibb. 2012. "Allocation of Communications to Reduce Mental Workload." *Procedia Computer Science* (8) (March): 242–248. https://doi.org/10.1016/j.procs.2012.01.052.

Pop, Vlad and W. Stuart Michelson. 2018. "Human Workload Modeling for Autonomous Ground Vehicles." In 2018 NDIA *Ground Vehicle Systems Engineering and Technology Symposium* 1–7. https://events.esd.org/wp-content/uploads/2018/07/Human-Workload-Modeling-for-Autonomous-Ground-Vehicles.pdf.

Powers, Dustin, and Andrew M. Gacy. 2018. "Using IMPRINT to Model Operator Staffing and Workload Considerations in a 24/7 Full-Service Mission Operations Center." In 2018 *SpaceOps Conference* 1–12. https://doi.org/10.2514/6.2018-2394.

Rusnock, Christina F., and Brett J. Borghetti. 2016. "Workload Profiles: A Continuous Measure of Mental Workload." *International Journal of Industrial Ergonomics* (63) (January): 49–64. https://doi.org/10.1016/j.ergon.2016.09.003.

Salvi, Lucia. 2001. Development of Improved Performance Research Integration Tool (IMPRINT) Performance Degradation Factors for the Air Warrior Program. Report No. ARL-TR-2311. Aberdeen Proving Ground, MD: Army Research Laboratory. https://apps.dtic.mil/sti/citations/ADA387840.

Sammer, Gebhard. 2006. "Workload and Electro-Encephalography." In *International Encyclopedia of Ergonomics and Human Factors*, edited by Waldemar Karwowski, 561–564. Boca Raton: CRC Press. https://doi-org.libproxy.nps.edu/10.1201/9780849375477.

Sargent, Robert. 2022. "3-Day IMPRINT Training at NPS." Presentation at Naval Postgraduate School, Monterey, CA, September 20–22, 2022.

Shivers, Matthew. 2021. "Holistic Situational Awareness and Decision Making (HSA-DM)." Presentation for the Naval Postgraduate School Human Systems Integration Department, Monterey, CA, November 2021.

Shivers, Matthew, and Rob Morony. 2021. "Human vs. Machine: Data for Situational Awareness, Decision Making and Robust Decisions." DEVCOM Aviation & Missile Center. April 20, 2021, https://rt.cto.mil/wp-content/uploads/T9_HFE-TAG-Shivers_Morony-Data-for-Human-vs-Machine-consumption-PAOrelease.pdf.

Stone, Peter, Gal A. Kaminka, Sarit Kraus, and Jeffrey S. Rosenschein. 2010. "Ad Hoc Autonomous Agent Teams: Collaboration without Pre-Coordination." In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (2010): 1504–1509*. https://dl.acm.org/doi/10.5555/2898607.2898847.

United States Army Aviation Center of Excellence. 2022. *UH-60 Series Aircrew Training Manual*. Fort Rucker, AL: Directorate of Training and Doctrine, Flight Training Branch.

Wickens, Christopher D. 2002. "Multiple Resources and Performance Prediction." *Theoretical Issues in Ergonomics Science* 3 (2) (November), 159–177. https://doi.org/10.1080/14639220210123806.

Wickens, Christopher D., and Amy L. Alexander. 2009. "Attentional Tunneling and Task Management in Synthetic Vision Displays." *The International Journal of Aviation Psychology* 19 (2) (April), 182–189. https://doi.org/10.1080/10508410902766549.

Young, Mark, and Neville Stanton. 2004. "Taking the Load Off: Investigations of How Adaptive Cruise Control Affects Mental Workload." *Ergonomics* 47 (9) (February): 1014–1035. https://doi.org/10.1080/00140130410001686348.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Fort Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

DUDLEY KNOX LIBRARY

NAVAL POSTGRADUATE SCHOOL

WWW.NPS.EDU

WHERE SCIENCE MEETS THE ART OF WARFARE