Assessing Cognitive Workload Prediction Models Using a Continuous Subjective Approach

Charles P. Rowan
Naval Postgraduate School
Monterey, California
charles.rowan@nps.edu

ABSTRACT

Adaptive automation has been identified as one of the most important topics in the history of human factors. Adaptive automation can help manage operator cognitive workload to acceptable levels to facilitate optimal performance. Future systems will become more complex and will have the potential to include more scalable autonomy as part of their design. Being able to assess cognitive workload in these future systems will be critical to ensure high levels of performance while mitigating negative outcomes for an operator. To investigate the impacts of future systems using these capabilities, this study leveraged cognitive workload prediction models using the Improved Performance Research Integration Tool (IMPRINT) to model operator workload while completing an adaptive automation scenario in NASA's Multi-Attribute Task Battery-II (MATB). After completing a task analysis, IMPRINT models were developed using the default anchors in IMPRINT and with feedback from expert users. Forty participants completed a 20-minute trial in MATB which consisted of multiple levels of workload and dynamically changing levels of automation. During completion of the task, the researcher prompted participants every 60 seconds to rate their experienced cognitive workload as a percentage of their maximum workload. This approach to subjective workload assessment is known as the Continuous Subjective Workload Assessment Graph (CSWAG) technique. CSWAG results from the study showed statistically significant differences between workload and automation conditions. Additionally, CSWAG results were correlated with the workload prediction models, serving to validate the CSWAG approach as a one method to assess the representational capacity of the IMPRINT models. This research provides an empirically based approach of coupling IMPRINT models with surrogate measures of cognitive workload. The resulting framework showed promise to gain insight more closely into an operator's experience with adaptive automation systems and can further be used to forecast workload impacts in future systems yet to be developed.

ABOUT THE AUTHOR

Charles Rowan is Lieutenant Colonel in the U.S. Army serving as an Assistant Professor and Interim Director of the Modeling, Virtual Environments, and Simulation (MOVES) Institute at the Naval Postgraduate School. He is an Army Functional Area 57 (Simulation Operations) Officer whose research areas focus on Human Systems Integration considerations in M&S. LTC Rowan was commissioned as armor officer from the U.S. Military Academy at West Point in 2005 and has served in a variety of leadership and staff positions in the 3rd Infantry and 1st Armored Divisions. He has served as an Assistant Professor in the Department of Behavioral Sciences and Leadership at West Point. LTC Rowan also served as the Division Simulations Operations Officer for the 2nd Infantry Division, and as the Chief of Knowledge Management for the 25th Infantry Division. LTC Rowan holds a B.S. in Engineering Psychology from West Point, and a M.S. in Human Systems Integration, and a Ph.D. in Modeling, Virtual Environments, and Simulation, both from the Naval Postgraduate School.

Validation of Adaptive Automation Cognitive Workload Prediction Models Using a Continuous Subjective Workload Assessment Graph (CSWAG) Approach

Charles P. Rowan
Naval Postgraduate School
Monterey, California
charles.rowan@nps.edu

INTRODUCTION

A primary objective of adaptive automation is to optimally manage operator workload (Parasuraman et al., 1992; Hilburn et al., 1993; Endsley, 2017). An assumption accompanying this objective is that levels of cognitive workload can be identified through surrogate measures to serve as a basis for changing automation. Adaptative automation has shown benefits by decreasing human error and performance variability (Scerbo, 1996). Adaptive automation can also enhance control of increasingly complex systems while mitigating operator performance variability, leading to reduced error rates (Scerbo, 1996; Woods, 1996). Additional intended outcomes of adaptive automation include keeping humans within a "band of proper workload" (de Greef & Arciszewski, 2009), assisting operator cognitive processes (Inagaki, 2003; Hancock et al., 2013; Kaber et al., 2006), and enhancing system performance (Brand & Schulte, 2017). These findings of reductions in cognitive workload and increases in performance suggest that the use of adaptive automation can serve as a viable intervention in high workload tasks (Endsley, 2017). Further, changes in operator performance should manifest at different levels of automation. As adaptive automation becomes more prevalent in increasingly complex systems, novel challenges are presented that need to be accounted for to address incongruence between the simulated system and real-world performance.

Adaptive automation can be triggered using three approaches: the critical-event strategy where high workload event times are determined ahead of time (Inagaki, 2003); the performance-measurement strategy where task performance is used to forecast states (Aricò et al., 2016; Inagaki, 2003); and the neurophysiological measurement strategy that uses various objective signals from an operator to determine correlates of workload (Scerbo et al., 2001). The present effort leveraged the critical-event strategy approach.

Task demands can drive cognitive workload and impact performance (Brand & Schulte, 2017; de Greef & Arciszewski, 2007; Hart, 2006; Inagaki, 2003; Kaber et al., 2001; Kaber & Endsley, 2004; Kanaan & Moacdieh, 2021; Smith & Baumann, 2020; Vagia et al., 2016). Given this relationship, workload measurement is an essential step to realize increased performance. Because cognitive workload is experienced subjectively by individuals, it can be described and assessed through introspection subjectively with reliability (Cain, 2007).

LITERATURE REVIEW

Cognitive Workload Subjective Measurement

There are numerous subjective methods to investigate cognitive workload, including the Malvern Capacity Estimate (MACE), Modified Cooper-Harper (MCH), Bedford Workload Rating Scale (BWRS), Subjective Workload Assessment Technique (SWAT), and Workload Profile (WP) (Vogl et al., 2020).

Another tool used for subjective cognitive workload assessment is the Continuous Subjective Workload Assessment Graph (CSWAG) (Miller & Shattuck, 2004). The CSWAG approach has participants report their experienced workload as a percentage of their maximum cognitive workload. Participants are asked to rate their cognitive workload percentage at given time intervals. These intervals must balance asking a participant their workload too often versus not enough to prevent introducing more cognitive workload or not capturing it at all.

Previous studies using CSWAG have asked participants to rate their percentage every minute during a task (Brown et al., 2021). Analogous approaches have been used with the SWAT at 30 second intervals and with the Instantaneous Self-Assessment (ISA) of workload at two-minute intervals (Brennan & Jordan, 1992; Zak et al., 2020). The CSWAG cognitive workload percentages can be generally categorized in three bins: lowest workload (0%-33%), just about right (34%-66%), and highest workload (67%-100%) (Miller & Shattuck, 2004).

The CSWAG was selected for this effort because of its ability to administer subjective workload assessments while minimizing disruptions to the primary task. Additionally, the use of more continuous inquiries into an assessed state at set intervals have shown sensitivity to changes in perceived cognitive workload (Brennan & Jordan, 1992; Brown et al., 2021; Zak et al., 2020).

The researcher instructs participants to provide a workload percentage when they hear "workload." The use of brevity in asking for the participants' workload and training them on the CSWAG in accordance with Figure 1 facilitates workload assessment during task completion. This approach helps mitigate disruptions to the operator's primary task to a negligible level.

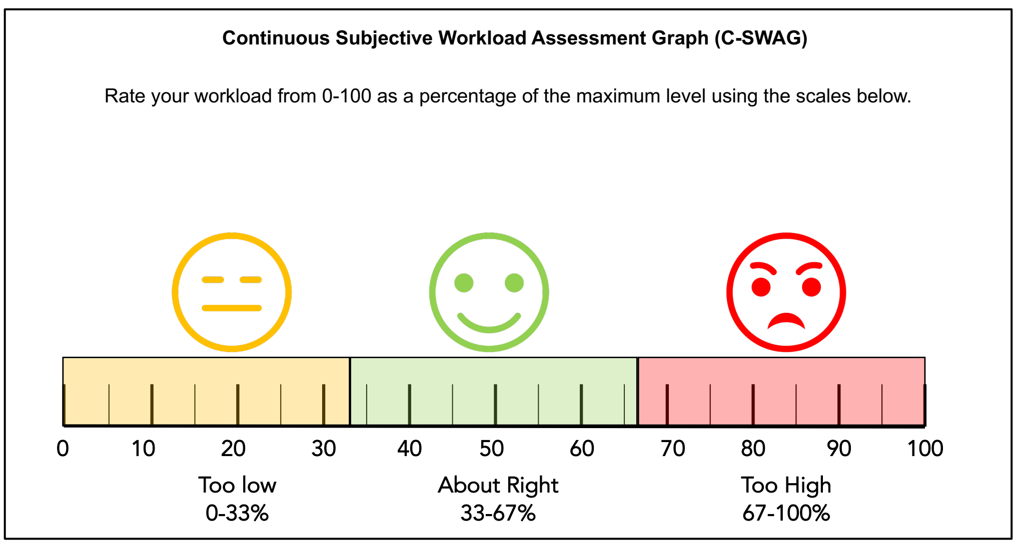

Figure 1: CSWAG participant reference visual.

Improved Performance Research Integration Tool (IMPRINT)

IMPRINT is a dynamic, stochastic, discrete event modelling and simulation tool that helps assess human interaction with a system and the resulting workload in completing a task or operation (Samms, 2010). In research applications, IMPRINT can provide a means to assess task analyses, workload modeling, and performance-shaping functions (Samms, 2010).

IMPRINT provides predictions of mental workload through both the Visual, Auditory, Cognitive, and Psychomotor (VACP) Theory (McCracken & Aldrich, 1984) and Multiple Resource Theory (MRT) (Wickens, 2002). A task analysis must be conducted first to decompose an operator's functions into tasks. A network of the task sequence is then developed. Once the initial task analysis is completed, a task network model can be simulated with relatively low overhead. IMPRINT can then run the modelled inputs to assess different factors in varying conditions.

Modelers using IMPRINT can link tasks with the required mental resources needed to accomplish them. They can then assign quantitative demand values based on the workload scale to each mental resource for the task, with descriptions for each demand level provided. Hardware and software components can be integrated into the model to demonstrate the human, machine, and environment representations are captured during a closed-loop cycle of an operation (Dahn & Laughery, 1997).

NASA's Multi-Attribute Task Battery II (MATB-II)

The NASA Multi-Attribute Task Battery II (MATB-II) is a computer-based simulation that allows for evaluation of human workload and performance (Santiago-Espada et al., 2011). MATB-II has flexibility in its configuration and execution through manipulation of its source files written in the Extensible Markup Language (XML). Numerous studies have leveraged MATB-II to study an operator's performance when executing multiple tasks in various domains. (Santiago-Espada et al., 2011). Recent studies have also modified the original version of MATB to include interfacing with ISA workload inputs, automated assistance and disruption, and the incorporation of live performance feedback to drive objectively and subjectively measured workload (Novstrup et al., 2023).

MATB-II includes four tasks presented through a user interface as seen in Figure 2: a system monitoring task (SYSMON), a tracking task (TRACK), a communications task (COMM), and a resource management task (RESMAN) (Santiago-Espada et al., 2011). MATB-II has a scheduling display to provide users communications and tracking task requirements within the next eight minutes. Fuel flow rates to support decision-making on the RESMAN task are displayed to the right of it. The MATB-II interface provides a Figure of Merit (FOM) beneath the scheduling display to inform the operator of their performance.

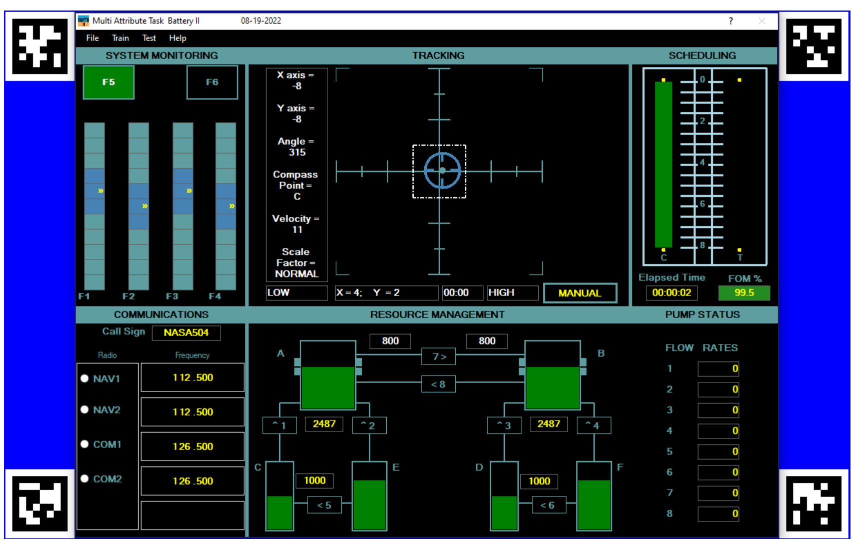

Figure 2. MATB-II user interface with April tags. Source: Santiago-Espada et al. (2011) and Olson (2010).

METHODS

IMPRINT Modelling

The researcher constructed four IMPRINT models that reflected the different scenarios presented to participants. These models reflected the accurate timing execution of the COMM and TRACK task. The researcher developed

IMPRINT models using the default value anchors organic to the system. The task network diagram for the MATB-II task used in the study is shown in Figure 3.

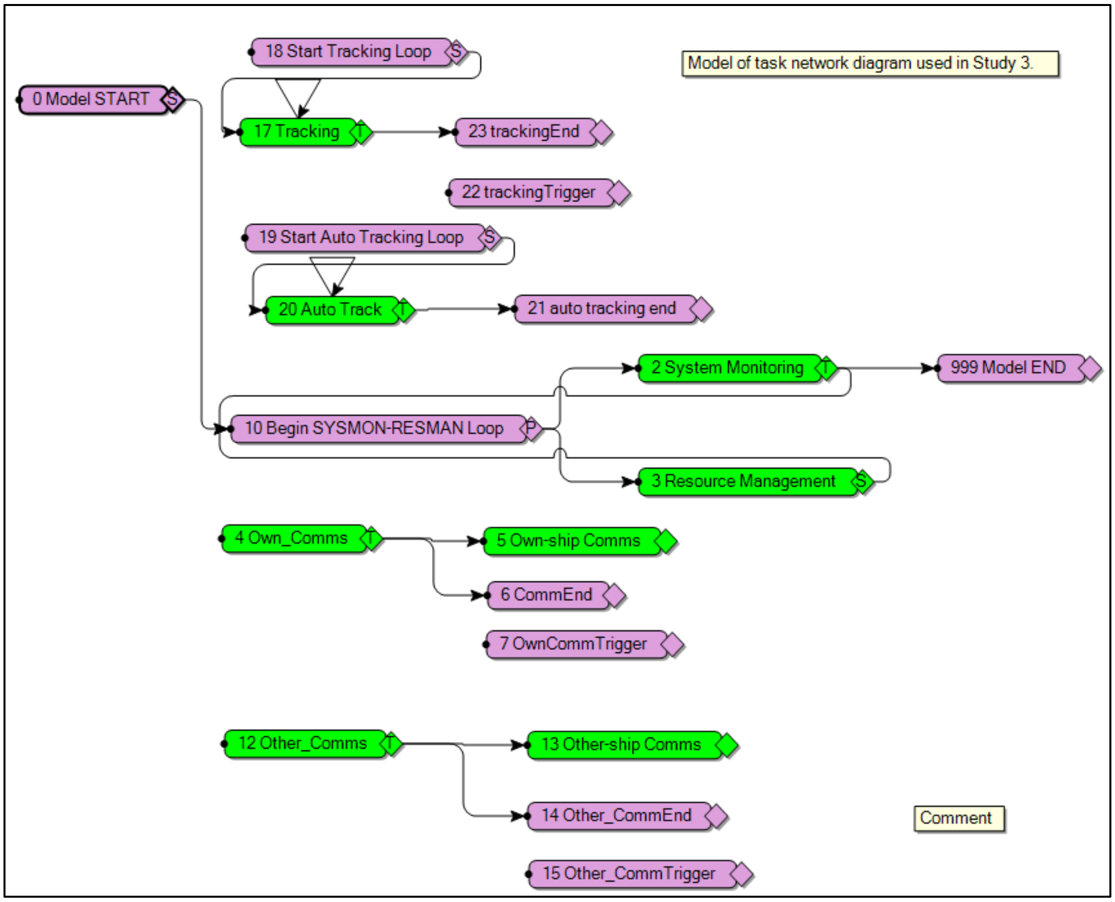

Figure 3. IMPRINT researcher-derived task network diagram.

The resource-interface demand values used in the study were chosen from the default values listed in IMPRINT. These IMPRINT-predicted time average workload values are listed in Table 1.

Table 1. Researcher-derived time-weighted predicted workload values.

Group	Low Manual	Low Auto	High Manual	High Auto	Total
1	40.55	11.70	41.54	25.25	26.51
2	40.31	14.34	43.48	21.12	27.41
3	40.20	11.81	42.35	24.29	28.37
4	39.95	14.92	41.29	20.61	28.15

Participants

The Naval Postgraduate School (NPS) Institutional Review Board (IRB) reviewed and approved the research methods used in this study. Participants were treated in accordance with the institution's human research protection program standards. All participants were informed of their rights as participants in the study and signed consent forms. Participants were recruited through personal communication, email, and campus-wide announcements.

Demographics

Forty-three participants were enrolled in the study, with 40 participants completing the study (mean age in years=34.18, SD= 4.90). Participants included 29 males and 11 females.

Materials

The two independent variables manipulated in this study were workload (low and high) and tracking condition (automated vs manual). Presentation of the workload and tracking condition were counterbalanced to account for order effects. The conditions used in the study are shown in Figure 4. Participants were randomly assigned to one of the four groups depicted in the figure. Participants interacted with MATB-II while seated, used a generic commercial joystick for the tracking task, and a mouse for inputs for the other three tasks.

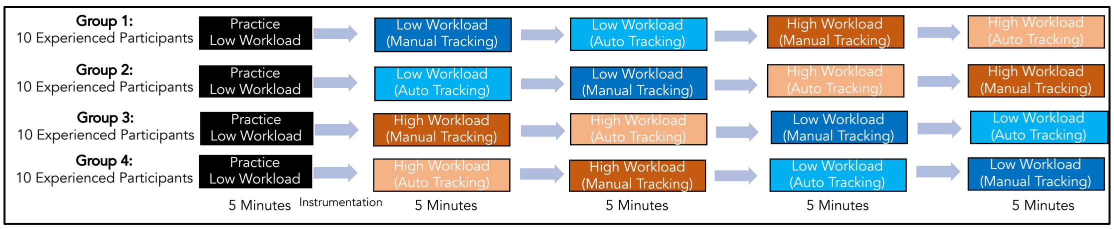

Figure 4. Experimental study conditions.

Participants were first trained on MATB-II through a video tutorial, part-task training sessions, four training sessions with all tasks that listed five minutes in length each. They then returned within 72-hours to go through the progression in Figure 4. The researcher assigned tasks throughout the scenarios in accordance with the parameters in Table 2. All participants were presented with both low and high workload conditions using McCurry et al.'s (2022) proposed task distribution.

Table 2. Study 3 MATB-II system settings for each condition.

	System Monitoring	Tracking	Communications	Resource Management
Low Workload (Manual Tracking)	11 Events	Low Joystick Response High Update Rate	3 Events	1 Pump Failure 1 Pump Shutoff
Low Workload (Auto Tracking)	11 Events	Automatic	3 Events	1 Pump Failure 1 Pump Shutoff
High Workload (Manual Tracking)	20 Events	Low Joystick Response High Update Rate	12 Events	10 Pump Failures 10 Pump Shutoffs
High Workload (Auto Tracking)	20 Events	Automatic	12 Events	10 Pump Failures 10 Pump Shutoffs

Prior to beginning training, participants were briefed on the CSWAG technique and provided the reference sheet with the image in Figure 1. Upon returning for their experimental trial, participants were refamiliarized with the CSWAG graphic and provided instructions on reporting their CSWAG percentage. The CSWAG reporting began 30 seconds after the trial started, and subsequent ratings were elicited every minute thereafter. The researcher would prompt the participant for their rating by saying "workload." Participants would respond verbally with their CSWAG assessment on the 0-100 percent scale. The researcher logged each response for later analysis. Additionally, CSWAG elicitations were delayed when a communications task was ongoing to eliminate any confounding auditory demands. Therefore, direct mappings to subjective workload experienced during communications tasks was difficult to ascertain. Figure 5 shows the experimental set up.

Figure 5. Experimental set up with MATB-II, physiological devices connected, and the CSWAG reference sheet above the monitor.

RESULTS

A mixed-effects model approach was used to analyze the data with JMP version 16.0.0. Fixed effects included workload level and automation condition. Participants were included as a random effect in the model and were nested within groups. There were no statistically significant differences between group conditions for CSWAG responses. These results indicate that there was no effect of presentation order between the groups, and therefore order is not included in the summary results listed in Table 3.

Table 3. Study summary results table.

Measure Category	**Measure Type**	**Dependent Measure**	**High vs. Low Workload**	**Tracking Condition**
Subjective Workload	CSWAG	Mean CSWAG	Low Workload -> lower CSWAG $\boldsymbol{p < .001}$	Lower CSWAG in Auto $\boldsymbol{p < .001}$

Mean CSWAG differences were statistically significant between both workload and tracking conditions as seen in Figure 6. Low workload resulted in lower reported CSWAG percentages than in the high workload condition (M=38.79, SD=14.13, SE=1.58 vs. M=48.75, SD=13.30, SE=1.49). Automatic tracking also resulted in lower reported CSWAG percentages than manual tracking (M=50.51, SD=12.37, SE=1.38 vs. M=37.04, SD=13.49, SE=1.51).

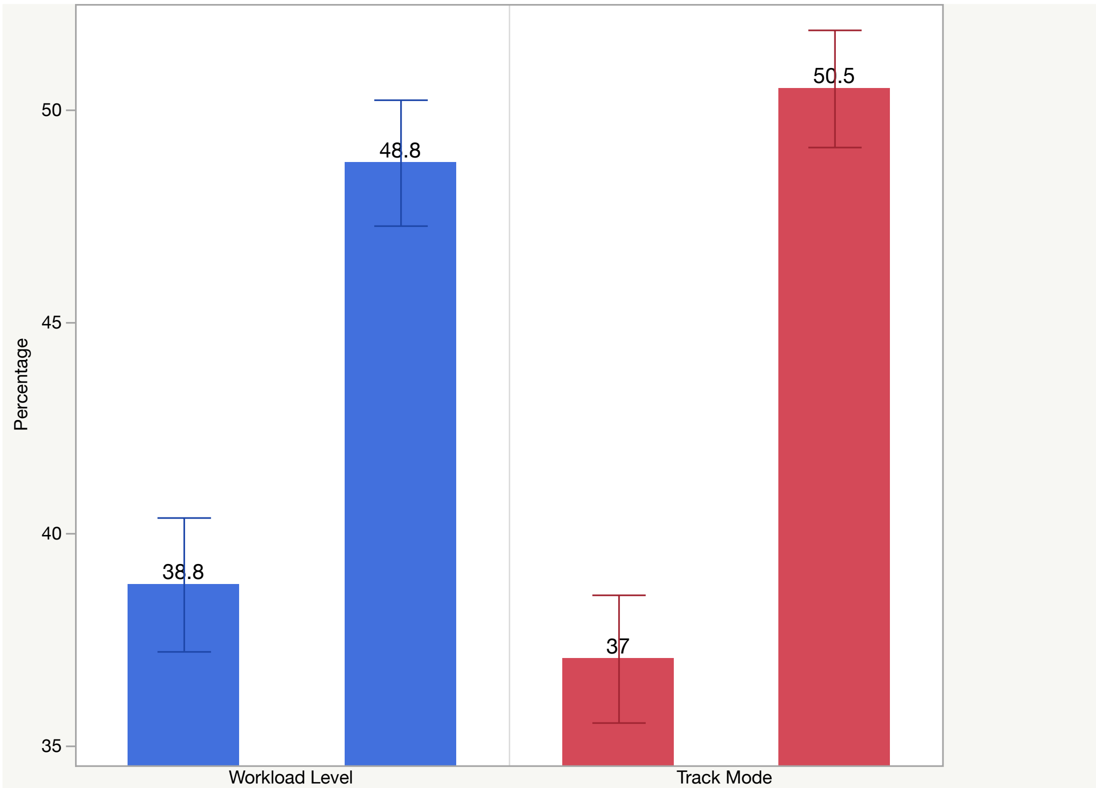

Figure 6. Study 3 CSWAG differences between workload (left) and tracking (right) conditions. Error bars denote the standard error.

DISCUSSION

The participants' reported CSWAG percentages were found to be sensitive to the different workload and adaptive automation conditions in the study. Composite results of all participants' CSWAG ratings are overlayed on the Group 1 workload prediction model in Figure 7. These results show the differences in workload conditions and overall pattern alignment with the IMPRINT workload prediction values.

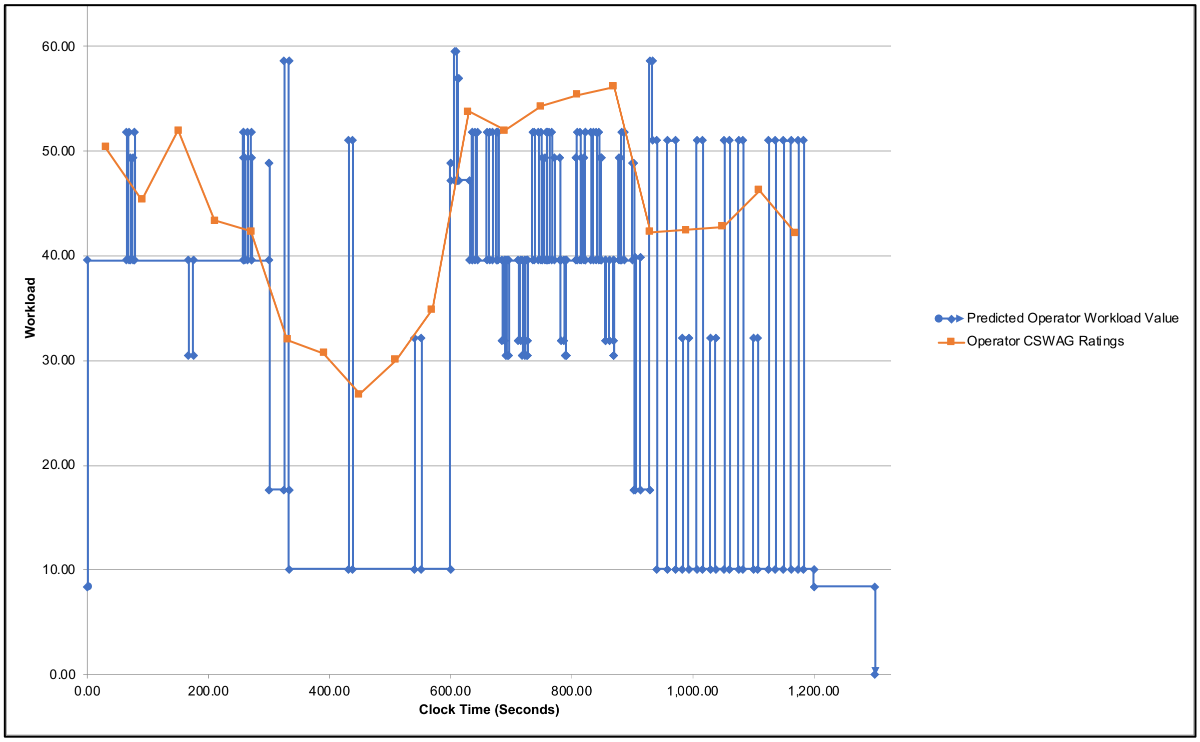

Figure 7. Study Group 1 IMPRINT Prediction Model with Composite CSWAG Values Overlayed.

These results serve as a validation of the IMPRINT prediction models using the CSWAG approach. Higher workload conditions models yielded higher workload prediction scores in IMPRINT. Conversely, lower workload condition models yielded lower workload prediction scores. This approach of using close-interval continuous subjective workload assessment suggests that workload changes are sensitive to experienced workload reports in an adaptive automation task.

CONCLUSION

This research highlighted the capability of a subjective workload assessment approach in CSWAG. However, this study had limitations for consideration. The experimental trials occurred in a control laboratory setting with a generally homogenous population. Additionally, the cognitive workload prediction models did not include aspects of sustained attention as a workload driver or the effects of fatigue. Future research could benefit from using CSWAG to gain insight into operator workload closer to real-time in more applied operational settings and in other domains that leverage AA. Multi-modal approaches to assessing cognitive workload are preferred to give more of a complete picture of the demands of a task (Cain, 2007). Combining subjective approaches like CSWAG with objective performance and psychophysiological data would also provide insight into the relationship between all three.

ACKNOWLEDGEMENTS

This paper is adapted from a portion of the author’s doctoral dissertation. The author would like to acknowledge the support from the U.S. Army’s Combat Capabilities Development Command’s Aviation and Missile Center Agency, and faculty of the Naval Postgraduate School who supported this research. This research is Distribution A, approved for public release; distribution is unlimited. The views expressed in this writing are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

REFERENCES

Brand, Y., & Schulte, A. (2017). Model-based prediction of workload for adaptive associate systems. *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 1722–1727.

Brennan, S. D., & Jordan, C. S. (1992). *Instantaneous Self-Assessment of Workload (ISA).*

Brown, A. J., Folger, J. R., Hardin, J. W., Moore, J. D., & Sica, Q. (2021). *Task Handoff Between Humans and Automation.* Naval Postgraduate School.

Cain, B. (2007). *A Review of the Mental Workload Literature.* Defence Research and Development Toronto (Canada).

Dahn, D., & Laughery, K. R. (1997). The integrated performance modeling environment—simulating human-system performance. *Proceedings of the 29th Conference on Winter Simulation*, 1141–1145.

de Greef, T., & Arciszewski, H. (2009). Triggering adaptive automation in naval command and control. In S. Cong (Ed.), *Frontiers in Adaptive Control*. InTech.

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human Factors*, 59(1), 5–27.

Hancock, P. A., & Matthews, G. (2019). Workload and performance: Associations, insensitivities, and dissociations. *Human Factors, 61*(3), 374–392.

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(9), 904–908.

Hilburn, B., Molloy, R., Wong, D., & Parasuraman, R. (1993). *Operator versus computer control of adaptive automation*. The Adaptive Function Allocation for Intelligent Cockpits (AFAIC) Program: Interim Research and Guidelines for the Application of Adaptive Automation, 31–36.

Inagaki, T. (2003). Adaptive automation: Sharing and trading of control. *Handbook of Cognitive Task Design*, 8, 147–169.

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science, 5*(2), 113–153.

Kaber, D. B., Perry, C. M., Segall, N., McClernon, C. K., & Prinzel III, L. J. (2006). Situation awareness implications of adaptive automation for information processing in an air traffic control-related task. *International Journal of Industrial Ergonomics, 36*(5), 447–462.

Kaber, D. B., Riley, J. M., Tan, K.-W., & Endsley, M. R. (2001). On the design of adaptive automation for complex systems. *International Journal of Cognitive Ergonomics, 5*(1), 37–57.

Kanaan, D., & Moacdieh, N. M. (2021). Eye tracking to evaluate the effects of interruptions and workload in a complex task. *Human Factors*.

McCracken, J. H., & Aldrich, T. B. (1984). *Analyses of Selected LHX Mission Functions: Implications for Operator Workload and System Automation Goals:* Defense Technical Information Center.

McCurry, C. D., Thanoon, M. I., & Zein-Sabatto, M. S. (2022). *Tennessee State University Multi-Attribute Task Battery (TSU-MATB)*. Department of Electrical and Computer Engineering, Tennessee State University.

Miller, N. L., & Shattuck, L. G. (2004). *A process model of situated cognition in military Command and control.* Naval Postgraduate School.

Novstrup, A., Tynan, M., Kline, J, De Luca, G., & Heaton, J. (2023). Towards Robust Estimation of Cognitive Workload from Wearable Physiological Sensors. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).*

Olson, E. (2010). AprilTag. https://april.eecs.umich.edu/software/apriltag

Parasuraman, R., Bahri, T., Deaton, J. E., Morrison, J. G., & Barnes, M. (1992). *Theory and Design of Adaptive Automation in Aviation Systems*. Naval Air Warfare Center.

Samms, C. (2010). Improved performance research integration tool (IMPRINT): Human performance modeling for improved system design. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 54*(7), 624–625.

Santiago-Espada, Y., Myer, R. R., Latorella, K. A., & Comstock, J. R. (2011). *The Multi-Attribute Task Battery II (MATB-II) Software for Human Performance and Workload Research: A User's Guide* (Technical Report NASA/TM–2011-217164). National Aeronautics and Space Administration.

Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. In *Automation and Human Performance: Theory and Applications*. CRC Press.

Smith, P., & Baumann, E. (2020). Human-automation teaming: Unintended consequences of automation on user performance. *2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC)*, 1–9.

Vagia, M., Transeth, A. A., & Fjerdingen, S. A. (2016). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? *Applied Ergonomics*, *53*, 190-202.

Vogl, J., Hayes, A., Aura, C., Milam, L., Temme, L., & St. Onge, P. (2020). *A Case for a Measurement of Incremental Workload.* U.S. Army Aeromedical Research Laboratory.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159–177.

Woods, D. D. (1996). Decomposing automation: Apparent simplicity, real complexity. In *Automation and Human Performance: Theory and Applications*. CRC Press.

Zak, Y., Parmet, Y., & Oron-Gilad, T. (2020). Subjective workload assessment technique (SWAT) in real time: Affordable methodology to continuously assess human operators' workload. *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2687–2694.