Naval Submarine Medical Research Laboratory

NSMRL/F2005/TR--2022-1407 October 31, 2022

A Cognitive Test Battery as a Performance Predictor of a Complex Task Requiring Sustained Attention: A Pilot Study

By

Chad Peltier
Sylvia Guillory
Jeffrey Bolkhovsky
David Gever
Margaret Wise
Krystina Diaz
Dawn Berger-Debrodt

Approved and Released by:
M. H. Jamerson, CAPT, MSC, USN
Commanding Officer
NAVSUBMEDRSCHLAB

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
31-10-2022	Technical Report	10/2020-10/2022

4. TITLE AND SUBTITLE

A Cognitive Test Battery as a Performance Predictor of a Complex Task Requiring Sustained Attention: A Pilot Study

5a. CONTRACT NUMBER

MIPR 11630649

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

'0603115HP

6. AUTHOR(S)

Chad Peltier, Sylvia Guillory, Jeffrey Bolkhovsky, David Gever, Margaret Wise, Krystina Diaz, Dawn Berger-Debrodt

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

F2005

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Submarine Medical Research Laboratory
Box 900
Groton, CT 06349-5900

8. PERFORMING ORGANIZATION REPORT NUMBER

NSMRL/F2005/TR--2022-1407

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

The Joint Program Committee-5/Military Operational Medicine Research Program (JPC-5/MOMRP)
1077 Patchel Street
Fort Detrick, MD 21702-5024

10. SPONSOR/MONITOR'S ACRONYM(S)

JPC-5

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The military relies on operators to efficiently identify threats to mission success and safety. Sonar operators search for threats, including hostile submarines, for extended periods, introducing the potential for a vigilance decrement, i.e., a decrease in performance over time. The U.S. Navy could benefit from a means to identify early which sonar operator candidates would be most likely to sustain high vigilance performance for prolonged periods. Currently, the Armed Services Vocational Aptitude Battery (ASVAB), a test similar to a standardized intelligence exam, is used to guide military Service Members' career tracks and specialties. However, the ASVAB may not adequately assess the vigilance or visual search abilities necessary for sonar or similar tasks, and a complementary assessment may improve upon the ASVAB's predictive validity for said tasks. This exploratory study investigated a 30-minute, 10-task cognitive test battery's ability to predict performance in a long duration, complex task, the Multi-Attribute Task Battery (MATB-II). The MATB-II demands sustained attention and simulates many of the demands of a submarine watch-station. MATB-II performance declined over time, confirming subjects' vigilance decrement. A regression using performance on the cognitive battery found that such performance accounted for 51% of the variance in overall MATB-II performance and 31% of the variance in sustained attention. This work suggests a 30-minute cognitive test battery may assist in identifying operators well-suited to performing comparable complex and attention-demanding vigilance tasks.

15. SUBJECT TERMS

individual differences, cognitive test battery, vigilance, sustained attention, multitasking

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			NAVSUBMEDRSCHLAB Commanding Officer
U	U	U	SAR	33	19b. TELEPHONE NUMBER *(Include area code)* 860-694-3263

Reset

[THIS PAGE INTENTIONALLY LEFT BLANK]

A Cognitive Test Battery as a Performance Predictor of a Complex Task Requiting Sustained Attention: A Pilot Study

Authors:
[1]Chad Peltier
[1,2]Sylvia Guillory
[2]Jeffrey Bolkhovsky
[1,2]David Gever
[1,2]Margaret Wise
[1,2]Krystina Diaz
[1,2]Dawn Berger-Debrodt

[1]Leidos, Inc.
[2]Naval Submarine Medical Research Laboratory

Approved and Released by:

JAMERSON.MATTHEW.HUNTER.1269308288
Digitally signed by JAMERSON.MATTHEW.HUNTER.1269308288
Date: 2022.10.31 15:50:17 -04'00'

CAPT M. H. Jamerson, MSC, USN
Commanding Officer
Naval Submarine Medical Research Laboratory
Submarine Base New London Box 900
Groton, CT 06349-5900

ADMINISTRATIVE INFORMATION

The views expressed in this report reflect the results of research conducted by the author and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government. This work was supported by funding work unit number F2005. The study protocol was approved by the Naval Submarine Medical Research Laboratory Institutional Review Board in compliance with all applicable Federal regulations governing the protection of human subjects, protocol number NSMRL.2021.0006. I am a military service member or employee of the U.S. Government. This work was prepared as part of my official duties. Title 17, U.S.C., §105 provides that copyright protection under this title is not available for any work of the U.S. Government. Title 17, U.S.C., §101 defines a U.S. Government work as work prepared by a military Service member or employee of the U.S. Government as part of that person's official duties.

[THIS PAGE INTENTIONALLY LEFT BLANK]

Abstract

The military relies on operators to efficiently identify threats to mission success and safety. Sonar operators search for threats, including hostile submarines, for extended periods, introducing the potential for a vigilance decrement, i.e., a decrease in performance over time. The U.S. Navy could benefit from a means to identify early which sonar operator candidates would be most likely to sustain high vigilance performance for prolonged periods. Currently, the Armed Services Vocational Aptitude Battery (ASVAB), a test similar to a standardized intelligence exam, is used to guide military Service Members' career tracks and specialties. However, the ASVAB may not adequately assess the vigilance or visual search abilities necessary for sonar or similar tasks, and a complementary assessment may improve upon the ASVAB's predictive validity for said tasks. This exploratory study investigated a 30-minute, 10-task cognitive test battery's ability to predict performance in a long duration, complex task, the Multi-Attribute Task Battery (MATB-II). The MATB-II demands sustained attention and simulates many of the demands of a submarine watch-station. MATB-II performance declined over time, confirming subjects' vigilance decrement. A regression using performance on the cognitive battery found that such performance accounted for 51% of the variance in overall MATB-II performance and 31% of the variance in sustained attention. This work suggests a 30-minute cognitive test battery may assist in identifying operators well-suited to performing comparable complex and attention-demanding vigilance tasks.

Keywords: individual differences, cognitive test battery, vigilance, sustained attention, multitasking

Author Note: Correspondence should be addressed to Jeffrey Bolkhovsky, PhD, Naval Submarine Medical Research Laboratory, Submarine Base New London, Groton, CT 06349-5900, email: jeffrey.b.bolkhovsky.civ@health.mil.

Table of Contents

Table of Tables

Table of Figures

Introduction

Several military tasks require operators' sustained attention over long periods of time to identify and prevent threats to service members, mission success, or equipment. These tasks include: sonar operation (which involves identifying adversaries prior to own ship detection), security screening (which involves identifying threats before they enter a protected area), or satellite image scanning (which involves identifying threats in static geographic images). The duration of an operator's shift in any of these critical tasks can be several hours. The ability to detect threats can decline in as few as 30 minutes,[1,2] however, raising concerns about the sustainment of performance and attention. Given the importance of quickly and accurately detecting targets or threats, developing methods to reduce operator failure is critical.

High aptitude in sustained attention and decision-making is a requirement for successfully performing tasks such as sonar monitoring; therefore, a screening method that assesses such cognitive abilities may help augment the current selection processes for these types of roles. The Armed Services Vocational Aptitude Battery (ASVAB), for example, is an assessment currently used by the U.S. military in determining suitability for occupation specialties. The ASVAB is similar to a standardized intelligence test, measuring mathematical and verbal academic achievement. In addition, the ASVAB measures technical knowledge content not found in standardized intelligence tests (i.e., electronics, mechanical, and auto/shop). Measures of these cognitive abilities show low or nonsignificant correlations with sustained attention performance.[3] Furthermore, the ASVAB score accounts for 16% of variance in overall military job performance[4] and 4% of variance in target detection performance in a visual search task.[5] Strengthening the relationship between Service Member occupation selection and possession of the cognitive abilities required to perform each occupation may improve the career success of Service Members and military readiness.

Identifying an ideal operator through an individual differences approach could be a more effective method of screening. This approach works by assessing observers on several traits predicted to be important for performing an unrelated task, then measuring how much variance in performance those traits account for in said unrelated task. This takes advantage of individual variation in cognitive traits by determining which traits are predictive of high task performance and who has high levels of those traits so that individuals' abilities are best utilized in the appropriate job (i.e., the duties in which they will be most likely to succeed). As the sustainment of performance is critical in many operational tasks, it is also crucial to assess how cognitive abilities, resulting performance, and attention are maintained or degrade over time. The vigilance decrement, a widely observed phenomenon defined by a decrease in target detection accuracy and/or an increase in reaction time as time-on-task increases, shows individual variations. Some people experience a performance decrement in as little as ten minutes, whereas others are able to resist declines in performance beyond 30 minutes.[6] While we know that vigilance performance varies widely among individuals, finding short-duration assessments to predict vigilance has been challenging. While some work has shown that cognitive flexibility (as assessed by the Wisconsin Card Sorting Task)[7] can predict vigilance, and it has been proposed that measures of working memory *should* predict vigilance, cognitive predictors of vigilance have been either sparsely investigated or sparsely reported.[8] Therefore, we are taking an exploratory approach to identifying individual differences in cognitive abilities that predict sustained attention by including measures that have been previously used successfully (Wisconsin Card Sort, working memory) alongside several other demanding tasks that may be related to sustained and/or overall

performance. Using this exploratory individual differences approach to examine the potential links among several cognitive abilities and vigilance may ultimately enable recruiters and instructors to select the people who already show aptitude in those cognitive domains, as well as inform targeted training to strengthen performance in those areas.

The individual differences approach has been executed in the past by administering a series of "predictor" tasks to participants, who also complete an "outcome" task.[3,9-13] The predictor tasks are often cognitive ability assessments, such as measures of working memory, sustained attention, visual search, or pattern matching. They may also include noncognitive characteristics, including personality. The "*outcome task,*" or the task of interest, is often one in which collecting data from the task is expensive. This can be because of the time-consuming nature of the task. A sustained attention task, by definition, takes a long time to complete. The same is true for tasks that involve low prevalence targets, where it takes time to present enough critical trials. Data collection can also be expensive because the equipment necessary to administer the real task is costly, such as a sonar monitoring or flight control. Predictor tasks are often fast, cheap, and easy to administer. A set of predictor tasks that account for a significant proportion of the outcome task's variance in performance can be identified in one sample of operators. Time and money can then be saved by administering only that set of predictor tasks to a new group of operators. Their outcome task performance can then be predicted using the values derived from the initial sample of operators. However, a predictive and generalizable model must first be established.

For example, to identify operators who excel at a baggage screening task, an ideal "outcome task" that simulates real-world baggage screening would involve characteristics such as: lengthy duration, rare targets (i.e., weapons), distractors, and time pressure (i.e., limited time to perform the task). Potential suitable predictor tasks would include a visual search task for a hard-to-find target. This type of predictor task can be quickly administered using a psychomotor vigilance task to measure sustained attention, a cognitive ability test, and a personality test. This set of predictors would be entered into a predictive model with performance in the outcome task as the criterion variable, which would allow researchers to identify which tasks are significantly predictive, and at what strength, of the outcome task. If strong predictors are identified, then the model encompassing this restrictive set of predictor tasks can be applied to a new sample of operators who only need to complete those predictor tasks, saving time and money by eliminating the need to measure performance on the actual outcome task (i.e., job performance criterion).

The current study applied the individual differences approach to investigate the use of a cognitive test battery in predicting performance of an operationally-relevant outcome task, towards enhancing the effectiveness of duty screening and assignment methods. Specifically, this study sought to identify those individuals who are able to sustain visual and auditory attention, task switch, and react quickly and accurately. The outcome task of interest used to measure these abilities is the Multi-Attribute Task Battery II (MATB-II),[14] a computer-based task where participants must simultaneously perform four different sub-tasks: system monitoring, resource management, tracking, and communications. It can be prolonged to tap into one's ability to sustain attention,[15] and it requires participants to quickly switch between performing different sub-tasks—an important ability in real-world duties. The MATB-II also has a wide variance in performance between participants.[16]

The set of predictor tasks was based on the cognitive test battery *Cognition*, a battery that covers a range of functions.[17] The tasks included in the battery assessed sensorimotor speed

(Motor Praxis Task [MPT]),[18] memory for complex figures (Visual Object Learning Test [VOLT]),[19] working memory capacity (Fractal 2-Back task [F2B]),[20] executive functioning (Wisconsin Card Sorting task [WCS]),[21] spatial orientation (Line Orientation Test [LOT]),[22,23] the ability to recognize emotions conveyed through facial expressions (Emotion Recognition Task [ERT]),[24,25] abstract reasoning and pattern recognition (Matrix Reasoning Test [MRT]),[18] complex screening and tracking (Digit-Symbol Substitution Task [DSST]),[26] risk taking behavior (Balloon Analog Risk Test [BART]),[27] and vigilance (Psychomotor Vigilance Task [PVT]).[28] The selection of these tasks provided abbreviated measures of well-established tests, allowing for a fast (read: inexpensive) measurement of several cognitive abilities. This battery was chosen because it includes measures of several cognitive abilities that have demonstrated success in predicting performance in unrelated outcome tasks. For example, working memory has been shown to predict visual search performance,[9,10] vigilance,[29] and reduced mind wandering while on task;[30] fluid intelligence has been shown to predict visual search performance;[11,29] the WCS has been shown to predict vigilance;[7] and vigilance itself has predicted visual search performance.[31,32]

This battery can also be conveniently administered on a tablet, with a short training period on how to perform each task, which permits the use of the cognitive test battery in real world settings. Finally, performance on this battery has been shown to predict performance in spaceflight tasks, with a real-world application requiring complex decision-making and fast reaction times.[33]

Participants completed the cognitive test battery followed by the MATB-II task. From the resulting performance data, two predictive models were developed: one for overall performance (representative of performance of all sub-tasks in the outcome task) and one for sustained performance (measured by the change in performance over time). This two-model approach may allow us to differentiate between predictors of performance and predictors of resistance to cognitive fatigue. Strengthening the relationship between Service Member occupation selection and correlation with the cognitive abilities required to perform particular military jobs might improve the career success of Service Members and military readiness.

Methods

Participants. Twenty-seven participants gave written informed consent to participate in this study and received monetary compensation after each session. The study was approved by the Naval Submarine Medical Research Laboratory (NSMRL) Institutional Review Board (IRB). Criteria for participation required normal or corrected-to-normal hearing and vision, including color vision. Recruitment was open and opportunistic and did not target any specific navy command or population. Three participants were excluded from analysis for having an incomplete session. The final sample included in the data analysis consisted of 24 participants; 12 were female (50%) and 12 were male (50%), and the group was between 19 and 64 years-old, with a mean age of 31.04 years (SD = 10.82).

Cognitive Predictor Tasks. The cognitive test battery (see Figure 1) consisted of ten tasks.[17] The 10 tasks were performed in the order listed below.[33]

Motor Praxis Task (MPT): This task involved clicking on randomly located squares that progressively decreased in size (an average of 4% on each trial). The task was comprised of 20 trials. Performance was measured by reaction time. MPT measured sensory motor speed.

Visual Object Learning Test (VOLT): In this task, participants memorized a set of 10 oddly shaped geometric figures (each presented for 5 seconds), before then being presented with 20 figures or trials. Participants had to determine which were new and which were old. Performance was measured by accuracy and reaction time. The task measured visual learning and spatial working memory.

Fractal 2-Back (F2B): Participants were presented with fractals in a sequence (1000 ms and an inter-stimulus interval of 1000 ms), and responded when the current stimulus matched the n-2 stimulus (stimulus presented two fractals ago). Participants viewed 60 images over two successive blocks. Performance was measured by accuracy (hits, misses) and reaction time. The task measured working memory capacity.

Wisconsin Card Sorting task (WCS): This task was to categorize cards on the basis of one of three dimensions: color, shape, or number of items on the card. The matching rule changed after a fixed number of trials, and trials timed out after 5 seconds. Participants completed 30 trials. Performance was measured by accuracy, number of perseveration trials (repeating the previously correct rule after the rule changed), and reaction time. The WCS task measured mental flexibility and abstraction.

Line Orientation Test (LOT): Participants were presented with two lines, one of which they rotated (via mouse clicks) to be parallel to the other. Difficulty was manipulated by varying the length of the two lines (shorter lines were 37% shorter in length) and how much the manipulated line rotated with each of the participant's manipulations (2° or 6°). There were 12 trials in the task. Accuracy was measured by the degrees between the target line's orientation and the manipulated line's orientation. Number of clicks was also reported. The LOT measured visual spatial processing and orientation.

Emotion Recognition Task (ERT): In this task, participants labeled a series of faces expressing different emotions, and classified each as happy, sad, angry, fearful, or no emotion.[34] The task involved 40 trials. Performance was measured with accuracy and reaction time. The ERT measured emotion recognition.

Matrix Reasoning Test (MRT): In 12 trials, participants picked a shape that completed a complex pattern. Performance was measured by accuracy and reaction time. The MRT measured complex reasoning.

Digit-Symbol Substitution Task (DSST): The participants were presented with pairs of numbers (1 to 9) and symbols, and then as quickly as possible had to pick the correct number that corresponded to the symbol presented elsewhere on the screen. Performance was measured by reaction time and the number of correct responses in 90 seconds. The DSST measured complex scanning and visual tracking, as well as cognitive functioning.

Balloon Analog Risk Test (BART): Participants pumped/inflated a balloon until they either collected a reward or the balloon popped. The reward amount ($) increased with each pump, as did the chance that the balloon popped. If the balloon popped, the participant would lose the reward. Thirty balloons were presented. Performance was measured by the average number of pumps the participant made. The BART measured risk taking.

Psychomotor Vigilance Test (PVT): Participants responded by pressing a button on a response box as quickly as possible during a 3-minute period when a stimulus was presented on screen. Performance was measured by reaction time and number of lapse trials. The PVT measured vigilant attention.

Outcome task: MATB-II. The MATB-II display presented two tasks on the top half of the screen and two tasks on the bottom half of the screen. In order from left to right and top to bottom, the tasks were: system monitoring (SYSMON), tracking (TRACK), communications (COMMS), and resource management (RESMAN) (see Figure 1). The scheduling panel (SCHED) allowed participants to "look ahead" eight minutes into the future at events that were to occur during the COMMS and TRACK tasks. The participants used a mouse and a joystick to make responses. In the SYSMON task, participants monitored two lights (buttons: one green, one red) and four scales consisting of moving bars, and had to respond whenever one of the lights changed color or when a moving bar went beyond a specific distance threshold. The frequency at which the two lights required a response and the deviations of the moving bars within the scales changed at an average rate of 6.5 seconds. The TRACK task required participants to use a joystick to keep a reticle centered within a small central target box. The reticle would often drift randomly away from the box when not manned by the operator in "Manual" mode. The frequency that the task was in "Manual" mode was, on average, every 35 seconds. The force that the reticle deviated was set to medium, as was the joystick sensitivity. The COMMS task required participants to listen for a specific call sign and respond to it by entering the correct frequency into one of four radios. Commands were announced approximately every 17 seconds, and on average, the participant needed to respond to a target command for every four distractors. The RESMAN task required participants to manipulate eight pumps to ensure that the fuel levels of two primary tanks (tanks A and B) remained within $\pm$500 of 2,500 units of fuel. The pumps transferred fuel from different reserve tanks, and would often temporarily break. The pumps failed every 6.5 seconds on average, and pump pairs 2 and 5 as well as 4 and 6 never broke simultaneously (shown in Figure 1).

Participants also completed the tasks UNRAVEL[35] and Letter Wheel[36] (see Figure 1) during the session that occurred on Day 1. The results from these tasks were not of interest in the current study.

Procedure. Participation took place across two separate days. On Day 1, participants were briefed and trained on the predictor and outcome tasks. The tasks were administered using a Lenovo ThinkPad Laptop (15.5 in monitor). For the cognitive test battery, participants completed an abbreviated version of seven of the ten tasks in the fixed order listed earlier. The software, E-prime (Psychology Software Tools Inc., Pittsburgh, PA, USA), controlled stimulus presentation. The VOLT, MRT, and BART did not have practice sessions and were not administered during Day 1 because of limits on the number of available images (however all

three tasks incorporated practice trials during their Day 2 administration). Practice sessions were intentionally limited. As the intended use-case for this cognitive assessment battery is to supplement screening a large number of military recruits, where the speed and expense of the screening is a priority, we wanted to measure the battery's effectiveness under conditions that more closely resembled real-world use, rather than bringing all subjects to their performance asymptote through time consuming practice sessions. After the cognitive test battery, participants completed the Letter Wheel and UNRAVEL tasks. For MATB-II, participants received written and verbal instructions from the researchers, then performed the sub-tasks for 5 minutes each (fixed order: SYSMON, TRACK, COMMS, RESMAN), followed by 30 minutes of performing all the tasks simultaneously.

On Day 2, participants completed the test session of the cognitive test battery followed by 60 minutes of the simultaneous MATB-II task. The average interval between Day 1 training and Day 2 testing was 10.38 ± 6.36 days. Depending on the task, responses were recorded via a Serial Response Box (Psychology Software Tools Inc., Pittsburg, PA, USA), the ThinkPad keyboard, a Dell wired optical mouse, or a Logitech joystick. At the end of both days,

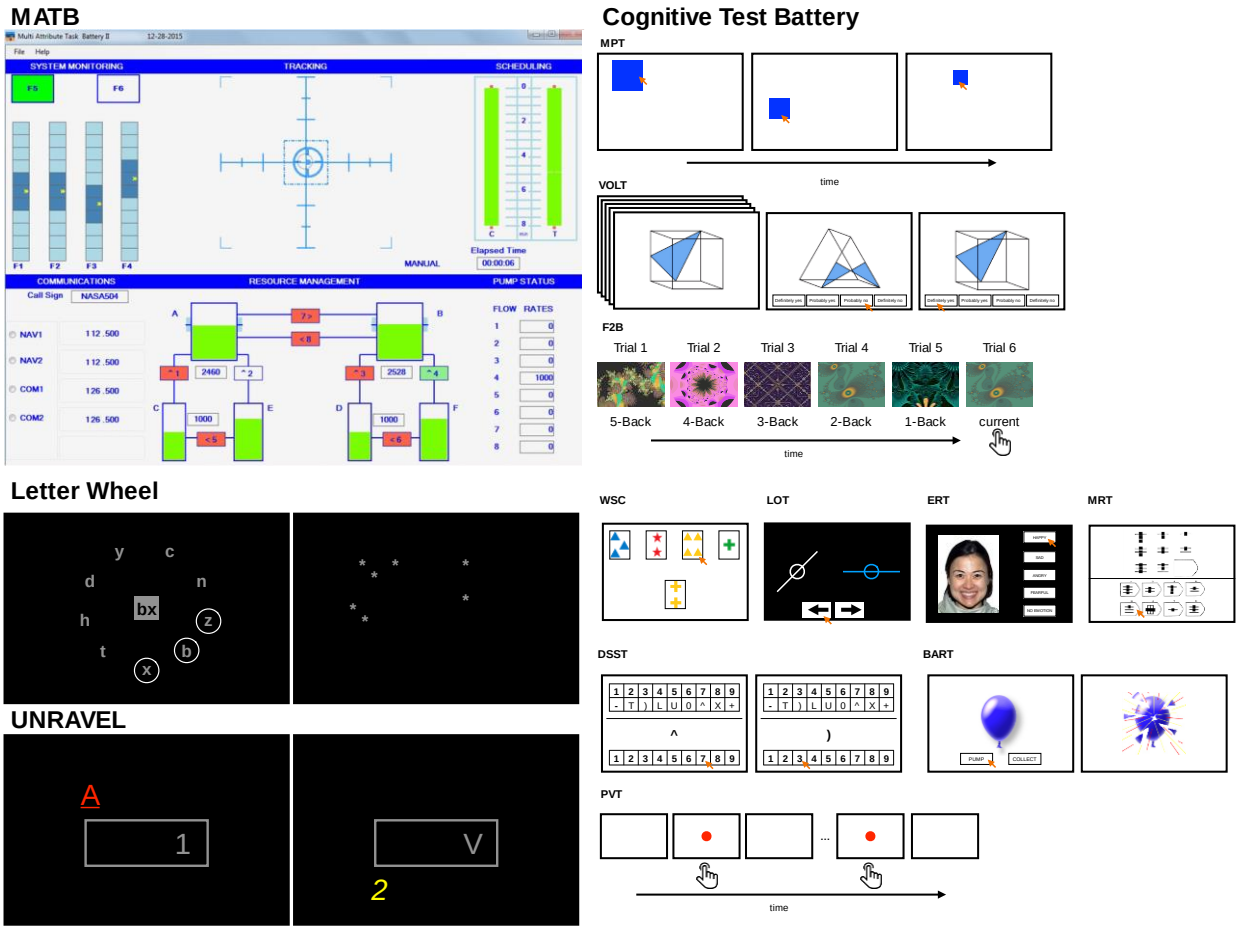

Figure 1. Experimental paradigms

MATB-II display (Top Left); example displays of the place-keeping tasks, Letter Wheel and UNRAVEL (Bottom Left); example displays and task schematic of each task in the Cognitive Test Battery (Right).

participants completed the National Aeronautics and Space Administration Task Load Index[37] (NASA-TLX) to estimate the workload required to perform the MATB-II.

Analysis

Correlations between the cognitive test battery scores during Day 1 (practice) and Day 2 (test) were conducted to examine the test-retest reliability of the test battery. Also, intraclass correlation coefficients (ICC) were obtained to assess the reliability of responses between the two sessions.

MATB-II results were analyzed in several ways. First, performance in each sub-task was calculated. For the SYSMON task, two dependent measures were calculated: the correct detection of the lights and scales ($Hits/(Hits + Misses)$) and reaction time. For the TRACK task, the average deviation of the reticle from the target box (calculated as root mean square error [RMSE]) was computed at 20 Hz every 15 seconds. For the COMMS task, both correct responses and reaction time were computed. The average deviations of the fuel tanks from the desired fuel level (difference score) were computed for the RESMAN task and summed across the two tanks.

An overall MATB-II composite performance score was then computed by applying the following method:[38,39] first, an average group score for each of the individual sub-tasks from Day 1 was calculated. Next, the group scores from the individual sub-tasks were used to normalize each participant's Day 2 sub-task scores, resulting in a z-score for each sub-task. These z-scores were then summed, and each task was given equal weight (i.e., one unit). Contributions from the reaction time and accuracy scores were adjusted by a factor of 0.5 for the SYSMON and COMMS to account for the two measures generated from these sub-tasks. The additive inverse of the reaction time (SYSMON and COMMS) and RMSE (TRACK sub-task) z-score measures were used to yield an opposite sign that was consistent with the accuracy measures. Higher scores were associated with higher accuracy. Likewise, the difference score measures from the RESMAN sub-task were computed as the inverse absolute value of the difference scores, in which higher scores were associated with better sub-task performance.

To investigate the change in outcome task performance as a function of time, in which a decrease in performance is interpreted as a time-on-task decrement, composite performance z-scores of each Day 2 (test session) sub-task were calculated for each 10-minute time interval across the entire MATB-II duration, amounting to six separate intervals or Blocks.

A linear mixed model was applied to model the association between time and MATB-II performance, with the participant as a covariance parameter. This analysis determined whether MATB-II performance changed over time. To obtain the change in performance for each individual, the MATB-II composite data were fitted to a linear function for each participant. The slope representing the magnitude of performance change was derived from the linear fit and was used as the response variable in a linear regression model.

A best-subset regression was used to identify the strongest predictors,[33] which included between 2 (high parsimonious model) and 10 (low parsimonious model) outcome measures from the cognitive test battery in predicting: (1) overall performance on the MATB-II task and (2) sustained performance throughout the MATB-II task (the magnitude of the performance change, i.e., the slope). A *jack-knife leave-one-out* cross-validation procedure evaluated the performance of the nine models containing the 2-10 predictors. In the *leave-one-out* cross-validation, one sample is left out as the test set (singleton) and the remaining samples form the training set used to predict the response value of the singleton, resulting in the mean square error calculation. This

process was repeated such that each sample is used once as a test set (n = 24), generating an average mean square error (MSE). The optimal model was determined as the model that: (1) had the highest proportion of significant predictors relative to the total predictors in the model and (2) minimized the prediction error (MSE). Correlations between the 20 cognitive test battery outcomes and both the composite and slope were also calculated.

The significance level was $p < .05$ and were not corrected for multiple comparisons; therefore, the results from this study should be considered exploratory.

Results

Cognitive Test Battery Test-Retest Reliability

Table 1 shows the correlations between the scores of the cognitive test battery on Day 1 (practice) and Day 2 (test). The mean correlation across measures was 0.13 (SD=0.29). The F2B accuracy and ERT reaction metrics showed the strongest correlation across sessions. Similarly, the ICC results aligned with the correlation analysis such that the ERT reaction time measure showed a moderate relatability between practice and test followed by the DSST and F2B measures.

Table 1. Test-Retest Correlations and ICC between Day 1 and Day 2 Cognitive Test Battery

	Correlation	ICC		
Cognitive Task	value, p-value	ICC	95% Confidence Interval	F-value, p-value
MPT: reaction time (ms)	r = 0.24, *p* = .26	0.15	[-0.27 0.51]	1.34, *p* = .24
F2B: accuracy (proportion hits)	r = 0.41, *p* = **.04***	0.31	[-0.10 0.63]	1.89, *p* = .07
F2B: reaction time (ms)	r = -0.34, *p* = .10	-0.002	[-0.40 0.40]	1.00, *p* = .50
WCS: accuracy (proportion correct)	r = -0.28, *p* = .19	-0.22	[-0.57 0.20]	0.64, *p* = .85
WCS: reaction time (ms)	r = 0.34, *p* = .10	0.29	[-0.12 0.61]	1.81, *p* = .08
WCS: # perseverative error trials	r = 0.21, *p* = .32	0.19	[-0.22 0.55]	1.47, *p* = .18
ERT: accuracy (proportion correct)	r = 0.21, *p* = .34	0.16	[-0.25 0.52]	1.37, *p* = .23
ERT: reaction time (ms)	r = 0.57, *p* = **.004***	0.56	[0.22 0.79]	3.59, *p* = **.002***
DSST: number of correct trials	r = 0.12, *p* = .59	0.006	[-0.39 0.40]	1.01, *p* = .49
DSST: reaction time (ms)	r = 0.38, *p* = .07	0.32	[-0.08 0.64]	1.95, *p* = .06
PVT: number of lapses	r = 0.22, *p* = .31	-0.06	[-0.45 0.34]	0.88, *p* = .62
PVT: reaction time (ms)	r = -0.16, *p* = .45	0.21	[-0.20 0.56]	1.53, *p* = .16

Significant correlations ($p<.05$) are indicated in bold with an asterisk (*)

Changes in MATB-II Performance across Time

The mean measure of performance for each MATB-II sub-task and overall score are provided in supplemental Table 1. The linear mixed model analysis revealed that MATB-II composite performance z-scores were significantly associated with elapsed time (overall model: $F(5,23) = 3.72$, $p = .013$). The individual effects of Blocks II-VI were all significantly lower than the initial Block I (Block II: $t(23) = -2.37$, $p = .026$; Block III: $t(23) = -3.14$, $p = .005$; Block IV: $t(23) = -2.80$, $p = .019$; Block V: $t(23) = -3.80$, $p = .001$; Block VI: $t(23) = -4.03$, $p = .001$). The estimate of the effects suggests an overall decline in performance from Block I to Block VI (relative mean changes from Block I: Block II: -1.82; Block III: -2.46; Block IV: -2.36; Block V: -3.05; Block VI: -3.38). Cognitive task mean performance is reported in supplemental Table 2.

Best Sub-Set Regression Model: MATB-II Overall Performance

The best fit model that minimized the MSE and significantly predicted MATB-II overall performance contained eight (MSE = 48.78; $F(8,15) = 3.97$, $p = .010$) predictors and accounted for 51% of the variance in MATB-II performance. The eight predictors were: BART mean number of pumps (B = -0.08, $p = .15$), ERT accuracy (B = -23.89, $p = .075$), F2B hits (B = -33.12, $p = .038$), F2B misses (B = -45.83, $p = .014$), F2B reaction time (B = -0.02, $p = .006$), LOT average error (B = 3.05, $p = .03$), MRT accuracy (B = -59.59, $p = .007$), and PVT reaction time (B = -0.04, $p = .014$). Results from cross-validation found a predicted $R^2 = 0.01$ and a ratio of the adjusted R^2 to R^2 of 0.74. The results of the best subset models that contained 2 through 10 cognitive task battery predictor variables are provided in supplemental Table 3.

Best Sub-Set Regression Model: MATB-II Sustained Performance

The best fit model that minimized the MSE and significantly predicted MATB-II sustained performance contained three predictors (MSE: 1.25; $F(3,20)=4.44$, $p = 0.015$) and accounted for 31% of the variance of the MATB-II slope. The three predictors were: WCS mean RT (B = -0.002, $p = .016$), BART mean number of pumps (B = 0.06, $p = .002$), and DSST number of correct trials (B = -0.05, $p = .047$). Results from cross-validation found a predicted $R^2 = 0.10$ and a ratio of the adjusted R^2 to R^2 of 0.76. The results of the best subset models that contained 2 through 10 cognitive task battery predictor variables are provided in supplemental Table 4.

Cognitive Battery and MATB-II Correlations

There were significant correlations between MATB-II overall score with the cognitive test battery performance measures, adjusting for age and sex (Table 2). There are reports on age[18] and sex difference[40] in cognitive abilities.[41] Specifically, males often outperform females on visual-spatial tasks,[42] and females generally perform better on social tasks, such as emotion recognition.[43] Overall, faster response speeds on the DSST and MRT were associated with higher MATB-II performance. Furthermore, higher accuracy on the WCS task was related to changes in performance with time on task. The MATB-II overall and sustained performance scores were not correlated ($r(24) = -0.086$, $p = .69$). There were no significant differences between sexes among the two response variables (MATB-II Overall: $t(22) = 0.-0.21$, $p = .84$, males = -3.41 ± 3.98, females = -3.12 ± 2.65; MATB-II slope: $t(22) = -0.11$, $p = .91$, males = -0.60 ±0.62, females = -0.57 ± 0.95; mean ± SD). Age was not correlated with either the MATB-II Overall score ($r(24) = -0.013$, $p = .95$) or MATB-II slope ($r(24) = -0.19$, $p = .38$). Correlations between cognitive outcomes are shown in Table 3.

Table 2. Correlations between Cognitive Test Battery Outcomes and MATB-II performance

	Bivariate		Partial adjusted for Age and Sex	
Task	MATB-II Overall	MATB-II Sustained	MATB-II Overall	MATB-II Sustained
BART avg. # of pumps	r=0.11, p=.61	r=0.052, p=.811	r=0.047, p=.836	r=0.008, p=.97
DSST # of correct trials	r=-0.057, p=.79	r=0.217, p=.308	r=-0.115, p=.612	r=0.19, p=.398
DSST reaction time	**r=-0.566, p=.004***	r=-0.049, p=.82	**r=-0.524, p=.012***	r=0.02, p=.929
ERT accuracy	r=-0.187, p=.382	r=0.305, p=.147	r=-0.235, p=.293	r=0.318, p=.149

ERT reaction time	r=-0.384, p=.064	r=0.056, p=.796	**r=-0.493, p=.02***	r=0.027, p=.906
F2B miss	r=0.185, p=.387	r=-0.118, p=.582	r=0.113, p=.616	r=-0.183, p=.414
F2B hits	r=0.188, p=.378	r=0.202, p=.343	r=0.147, p=.514	r=0.178, p=.428
F2B reaction time	r=-0.175, p=.412	r=-0.077, p=.72	r=-0.059, p=.795	r=0.007, p=.976
LOT # of clicks	r=-0.158, p=.462	r=0.058, p=.786	r=-0.048, p=.833	r=0.155, p=.49
LOT avg. errors	r=-0.36, p=.084	r=-0.052, p=.81	r=-0.266, p=.232	r=0.059, p=.793
MPT reaction time	r=-0.061, p=.776	r=0.27, p=.201	r=-0.25, p=.261	r=0.204, p=.363
MRT reaction time	**r=-0.516, p=.01***	r=-0.068, p=.753	**r=-0.56, p=.007***	r=-0.062, p=.784
MRT accuracy	r=-0.196, p=.359	r=-0.142, p=.507	r=-0.109, p=.628	r=-0.086, p=.705
PVT # of lapses	r=-0.244, p=.25	r=0.01, p=.964	r=-0.286, p=.197	r=0.012, p=.957
PVT reaction time	r=0.014, p=.948	r=0.092, p=.67	r=0.003, p=.991	r=0.088, p=.698
VOLT accuracy	r=-0.212, p=.32	r=0.018, p=.932	r=-0.367, p=.093	r=-0.05, p=.826
VOLT reaction time	r=-0.013, p=.95	r=0.231, p=.277	r=0.032, p=.888	r=0.27, p=.224
WCS accuracy	r=0.054, p=.802	**r=0.436, p=.033***	r=0.027, p=.905	**r=0.427, p=.047***
WCS perseverate error	r=0.33, p=.116	r=0.031, p=.886	r=0.241, p=.279	r=-0.062, p=.784
WCS reaction time	r=0.148, p=.489	r=0.109, p=.612	r=0.012, p=.959	r=0.022, p=.924

Significant correlations ($p<.05$) are indicated in bold with an asterisk (*)

Table 3. Correlations between cognitive test battery task outcomes, r coefficient and p-value (p-values in parenthesis)

Task	DSST # of trials	DSST RT	ERT acc	ERT RT	F2B miss	F2B hits	F2B RT	LOT # of click	LOT avg. error	MPT RT	MRT RT	MRT acc	PVT # of lapses	PVT RT	VOLT acc	VOLT RT	WCS acc	WCS error	WCS RT
BART # pumps	-0.20 (.34)	0.24 (.26)	-0.04 (.86)	-0.12 (.58)	0.04 (.87)	-0.13 (.56)	-0.24 (.26)	-0.12 (.58)	0.03 (.90)	-0.17 (.44)	-0.12 (.58)	0.25 (.24)	-0.32 (.13)	-0.32 (.13)	0.06 (.78)	-0.05 (.80)	-0.003 (.99)	0.25 (.25)	0.40 (.05)
DSST # of trials	------	-0.98 **(<.001*)**	-0.28 (.18)	0.60 **(.002*)**	0.44 **(.03*)**	-0.38 (.07)	0.41 **(.05*)**	-0.31 (.14)	0.02 (.92)	0.71 **(<.001*)**	0.42 **(.04*)**	-0.09 (.69)	0.06 (.77)	0.25 (.24)	0.28 (.19)	-0.18 (.41)	-0.14 (.51)	-0.37 (.07)	0.55 **(.01*)**
DSST reaction time		------	0.29 (.16)	-0.62 **(.001*)**	-0.43 **(.04*)**	0.38 (.07)	-0.40 (.05)	0.27 (.20)	-0.05 (.83)	-0.71 **(<.001*)**	-0.43 **(.04*)**	0.09 (.68)	-0.11 (.62)	-0.30 (.15)	-0.31 (.14)	0.15 (.48)	0.12 (.59)	0.37 (.08)	-0.52 **(.01*)**
ERT accuracy			------	-0.52 **(.009*)**	-0.40 (.05)	0.35 (.10)	-0.14 (.51)	-0.28 (.19)	0.28 (.18)	-0.19 (.37)	-0.16 (.46)	0.05 (.81)	-0.05 (.83)	-0.27 (.21)	-0.10 (.64)	0.17 (.44)	0.35 (.10)	0.00 (1.00)	-0.38 (.07)
ERT reaction time				------	0.14 (.51)	-0.09 (.67)	0.26 (.23)	-0.11 (.62)	-0.03 (.90)	0.52 **(.01*)**	0.52 **(.01*)**	-0.06 (.78)	0.01 (.96)	0.17 (.43)	0.45 **(.03*)**	-0.04 (.87)	-0.002 (.99)	-0.31 (.14)	0.38 (.06)
F2B miss					------	-0.95 **(<.001*)**	-0.12 (.56)	-0.17 (.44)	-0.17 (.44)	0.36 (.09)	0.36 (.08)	-0.08 (.72)	-0.03 (.89)	0.07 (.76)	-0.20 (.35)	-0.12 (.57)	-0.36 (.08)	-0.08 (.70)	0.20 (.36)
F2B hits						------	0.18 (.39)	0.16 (.47)	0.2 (.35)	-0.27 (.21)	-0.29 (.16)	0.04 (.86)	-0.03 (.90)	-0.08 (.73)	0.21 (.32)	0.02 (.92)	0.27 (.20)	0.10 (.65)	-0.19 (.37)
F2B reaction time							------	0.02 (.92)	0.38 (.07)	0.36 (.09)	0.34 (.10)	-0.09 (.68)	0.35 (.09)	0.41 **(.05*)**	0.30 (.16)	-0.24 (.26)	0.13 (.55)	-0.22 (.31)	0.20 (.34)
LOT # of clicks								------	-0.24 (.25)	-0.31 (.14)	-0.10 (.66)	0.14 (.52)	-0.07 (.75)	-0.02 (.93)	-0.10 (.64)	0.15 (.48)	0.24 (.26)	0.21 (.31)	-0.36 (.09)
LOT avg. errors									------	0.20 (.36)	0.29 (.16)	0.24 (.26)	0.04 (.84)	0.09 (.69)	0.20 (.34)	0.16 (.45)	-0.02 (.92)	-0.09 (.69)	0.15 (.50)
MPT reaction time										------	0.33 (.12)	0.26 (.23)	-0.13 (.56)	0.16 (.45)	0.13 (.55)	-0.05 (.82)	-0.06 (.78)	-0.23 (.29)	0.5 (.01*)
MRT reaction time											------	0.03 (.90)	0.23 (.28)	0.25 (.23)	-0.03 (.89)	0.19 (.39)	0.20 (.35)	-0.20 (.34)	0.05 (.82)
MRT accuracy												------	-0.48 **(.02*)**	-0.37 (.07)	-0.10 (.66)	0.51 (.01)	0.30 (.16)	0.34 (.11)	0.09 (.67)
PVT # of lapses													------	0.88 **(<.001*)**	-0.09 (.67)	-0.26 (.22)	-0.07 (.76)	-0.39 (.06)	-0.13 (.55)
PVT reaction time														------	-0.01 (.96)	-0.22 (.31)	-0.13 (.54)	-0.45 **(.03*)**	0.02 (.92)
VOLT accuracy															------	-0.18 (.39)	0.03 (.91)	-0.14 (.53)	0.15 (.49)
VOLT reaction time																------	0.38 (.06)	0.17 (.44)	-0.18 (.40)
WCS accuracy																	------	0.19 (.38)	-0.45 **(.03*)**
WCS perseverate error																		------	-0.20 (.36)

Significant correlations (p<.05) are indicated in bold with an asterisk (*)

Discussion

Twenty-four participants completed two days of testing; the first day was devoted to practicing the MATB-II task and the cognitive test battery, and the second day involved testing on each of these tasks. The critical outcome measure, MATB-II, showed a significant decline in performance over time, as indicated by the significant changes in composite performance z-scores over elapsed time (Blocks I to VI). Given two different measures of performance in MATB-II, overall performance and sustained performance (slope), an individual differences approach was used to determine whether a cognitive test battery accounted for a significant proportion of the variance in MATB-II performance. A best-subset regression indicated that BART average number of pumps, ERT accuracy, F2B hits, F2B misses, F2B reaction time, LOT average errors, MRT accuracy, and PVT reaction time accounted for 51% of the variance in overall performance. The same method revealed that WCS reaction time, BART mean pumps, and DSST number of accurate trials accounted for 31% of the variance in MATB-II sustained performance.

The predictors of overall performance involved measures from the BART, ERT, F2B, LOT, MRT, and PVT. The PVT was expected to be a robust predictor of performance on a long duration task, as the PVT has long been established as a functional measure of the consequences of fatigue.[44] The regression results indicated that for an increase in reaction time on the PVT there was a significant decrease in overall MATB-II performance. This is similar to a previous finding that showed poor performance on a vigilance task correlated with poor performance in a visual search task.[3,29] Risk propensity measured on the BART indicates that a willingness to make risky decisions is disadvantageous during tasks similar to MATB-II. This may have been relevant during the resource management sub-task, when participants decided which pumps to (de)activate to maintain adequate levels of fuel. It may also apply on a broader level when participants had to decide to which sub-tasks they would dedicate their cognitive resources. Event response speed and accuracy related to working memory (F2B) and complex reasoning (MRT) are all linked to multitasking performance.[45,46] The inclusion of these cognitive measures in the best-fit model emphasizes the importance of these skills in a multitasking environment such as MATB-II, and also aligns with previous research showing strong relationships between working memory, fluid intelligence, and a demanding yet unrelated task.[3,29] Interestingly, there was a relationship between accuracy of emotion recognition (ERT) and MATB-II total performance. This result aligns with research finding that measures of emotion recognition were correlated with performance on a simulated space docking task.[33] The results of the current investigation also highlight the importance of spatial orientation (LOT) for complex layouts such as the MATB-II task, which featured multiple control panels. The LOT may be capturing the contribution of spatial orientation of the MATB-II task.

In addition to the benefits that an individual differences analysis has over using a cognitive ability test like the ASVAB as the sole predictor of overall job performance, the regression of sustained performance showed that the WCS reaction time, BART mean pumps, and DSST number of accurate trials accounted for approximately a third of the variance in performance change over time. This battery successfully accounting for a significant portion of the variance in performance change over time suggests that when screening for roles where sustained attention may be important for job performance, such as sonar, security screening, or flight control, the military should consider incorporating these or similar assessments into their

existing screening process to better identify those who are best suited for the role. The DSST is a widely used measure of information processing speed and is featured as a subtest in the Wechsler Adult Intelligence Scale (WAIS-IV), a test designed to measure intelligence and cognitive ability.[47] High performance on the DSST involves associative learning, motor speed, working memory, and visual tracking. Although it is not surprising that a task related to an intelligence scale predicts task performance, our finding that WCS and BART account for variance over and above this proxy measure of intelligence shows the advantage and potential utility of this cognitive assessment of individual differences. Finding that the WCS predicts vigilance performance replicates previous research, increasing our confidence in the robustness of this result.[7] In the military, candidates are screened for jobs using the ASVAB, and those with high scores are recommended into more cognitively demanding roles. However, the ASVAB is limited, as it is mostly a measure of crystalized intelligence,[48] and only accounts for approximately 16% of the variance in military job performance[49] and approximately 4% of variance in target detection performance in visual search tasks.[5] With the WCS and BART performance accounting for variance above that accounted for by a DSST, a component used to measure intelligence, the individual differences method could offer an opportunity to improve the predictive validity of military screening for sustained attention job performance.

While the results of the present study are promising, datasets with modest sample sizes may be prone to overfitting. In an attempt to overcome this, this study applied a leave-one-out cross-validation technique. However, predicted R^2 values, a measure of overfitting,[50,51] for both overall MATB performance and sustained attention were less than 0.1, indicating that the selected optimal regression models may not predict responses to new observations well. A second metric used to determine model fit when predicting new data, the ratio of the adjusted R^2 to R^2,[51] found values that were less than .80. A ratio of 1.0 would indicate that the models overfit the existing data. The selected models, thus, demonstrated some moderate levels of overfitting. Therefore, results from the best-subset analyses should be interpreted with some caution.

While the cognitive test battery showed low test-retest reliability, these results are in line with prior research using a similar test battery.[17,41] These results could be due to the modest sample size of this study and that the practice session (Day 1) was used for the analysis. Indeed, studies have found the greatest variability between initial and second sessions of the cognitive test battery.[17] The establishment of normative data with larger sample sizes for defined populations would be beneficial. There was also some degree of correlation among the cognitive test battery outcome measures. This would suggest that some of the predictors could potentially be removed, such as the DSST, as it significantly correlates with five of the ten tests. By eliminating some tasks from the battery, the remaining tasks with the best unique prediction for overall and sustained attention performance would be left in the models.

Beyond the benefits that the individual differences approach has in predicting task performance over measures primarily of intelligence, it may also be inexpensive and fast to administer. The cognitive test battery investigated in the current study takes approximately 30 minutes to complete (even after accounting for training time, which may not be necessary) and can be administered on a portable tablet.

Using the individual differences approach as a targeted training tool to optimize performance has several additional advantages over methods that have focused on experimental measures to improve performance during sustained attention tasks post-duty selection. These measures include experimental manipulations that may be limited in achieving performance improvements.[52-54] Of the six different methods investigated by Wolfe et al 2007, only one

involving feedback to the observer was found to improve target detection rates during a sustained visual search task, although it simultaneously increased false alarms. Increasing target detections while simultaneously increasing false alarms is suboptimal because a false alarm, especially in a military context, can lead to targeting the wrong person, area, or ship. Similarly, increasing the perceived target prevalence by adding false feedback about "missed" targets improved threat detection rates, but this also increased false alarms.[55] Peltier and Becker (2017) and Drew and Williams (2017) both used similar methods in an attempt to improve visual search performance. In an effort to guide the participant's attention to unsearched areas of the display, they visually marked the areas of a search display that observers fixated on. However, this did not improve performance in nine out of ten experiments. Even in methods where target detection accuracy was increased by engaging observers in a similarity search (observers were instructed to find the element in the display most similar to a predefined target, rather than a more typical target present/absent response[56]), implementation into real-world, real-time dynamic visual displays is impractical and does not provide for the break in search where in the search array disappears for making such responses. Overall, experimental methods have failed to reliably increase target detection rates without a serious cost (e.g., increased false alarms), and would be too impractical, too difficult, and too costly (e.g., via equipment modifications) to implement, even if effective. It makes sense to identify the people who express a natural ability to succeed at sustained attention tasks by applying the individual differences approach, instead of applying performance improvement strategies once individuals have already been assigned to these tasks.

Limitations

Though the initial results presented in the current study are a promising indicator that a cognitive test battery may be able to predict operational performance, there are several limitations to this work that should be addressed in future research. First, the outcome measure, while complex and difficult, was not directly representative of the operational task; thus, the ability of the cognitive battery to truly predict operational performance is unknown. Future research should assess the ability of the cognitive battery to predict performance in a real-world task, such as sonar monitoring. Similarly, the duration of the operational task should be increased beyond the one hour used here, as decrements in performance may change over increasingly extended periods of time, potentially changing the accuracy of the cognitive battery in predicting the degree of the vigilance decrement. Second, this exploratory work relied on a relatively small number of participants, especially given the high number of predictor tasks. Future research should use a broader participant pool, providing additional statistical power to the results. Third, although we have approached this problem from an applied, military perspective, evaluating whether the use of a cognitive screening task can improve the performance of the military, the mechanisms behind the relationships between our predictors are unclear. Future research with a more academic focus may be able to elucidate these relationships between predictor and outcome. Fourth, given that the goal is to augment military job selection and screening, this work could be improved if the cognitive battery's predictive validity was compared to that of standard screening tools like the ASVAB. Potentially, such work could suggest that the two tools account for unique variance, in which case both tools would have utility. A hybrid screening approach would potentially be appropriate. Fifth, because we only conducted one practice session for the cognitive battery tasks, which likely prevented observers' performance on each task from reaching asymptote, our model may have inaccurate estimates of observer's abilities, limiting the accuracy of the model. Though this is a limitation, we would also like to note that the real-world

use of this method would likely have zero or one practice sessions (to save time and money when using this method on thousands of military recruits), so its use here may reflect real-world conditions.

While the cognitive test battery showed poor to moderate test-retest reliability, this observation may have potentially limited effects on the subsequent analysis and generalizability of the conclusions. This is because, first, the reliability calculations were based on a practice session. Large individual variations may be attributable to learning during the practice session.[57] Second, the test battery is derived from the well-established and validated Penn Computerized Neurocognitive battery.[25,40,41] Nonetheless, future research should still continue to re-confirm the relatability of these tests, especially as the software expands to newer platforms such as tablets and iPhones.

Future work must also consider the combination of participant pool and task. As noted, future research should investigate the effectiveness of individual differences-based screening tools to predict performance in more realistic, duty-representative tasks, but it must also consider the use of expert vs. novice participants. Whereas the generalizability of the model may be maximized through a real-world task performed by a trained participant pool (i.e., sonar operators completing a sonar monitoring task), the model may still be effective and generalizable with a novice pool of participants, such as the one used here. Stakeholders in screening or training programs may believe that any inherent individual differences in complex task or sustained attention performance may be overcome through training, minimizing the utility of the approach described here. However, research shows that inherent individual differences still account for performance differences after thousands of hours of training.[58,59] Such work supports the approach of using screening to identify those who are particularly well suited for a role using an individual differences approach, before training those with aptitude to become proficient, real-world performers. Future work should also consider non-cognitive attributes in predicting sustained attention performance. Research has found that motivational and personality factors influence performance.[60] Capturing both cognitive and non-cognitive attributes may help to improve the effectiveness of the military job selection process.

Conclusion

Despite these concerns that must be addressed in future research, our work suggests that it is possible to account for 51% of the variance in one's ability to perform complex tasks, and 31% of the variance in sustained performance, using a convenient and cheap to administer cognitive battery. We are not suggesting a cognitive test battery should replace or augment the ASVAB in military screening, as its predictive validity for actual job performance is yet to be determined. Future research is needed to further examine the cognitive tasks to ensure their reliability by showing repeatable measures of individual performance across multiple days. However, its advantages as initial indices of predictive validity for a difficult task and predictive validity for sustained attention, suggest that its potential for use (standalone or supplementary) is worth further investigation.

References

1. Pattyn N, Neyt X, Henderickx D, Soetens E. Psychophysiological investigation of vigilance decrement: boredom or cognitive fatigue? *Physiol Behav.* 2008;93(1-2):369-378.
2. Nuechterlein KH, Parasuraman R, Jiang Q. Visual sustained attention: image degradation produces rapid sensitivity decrement over time. *Science.* 1983;220(4594):327-329.
3. Peltier C, Becker MW. Individual differences predict low prevalence visual search performance and sources of errors: An eye-tracking study. *J Exp Psychol Appl.* 2020;26(4):646-658.
4. Ree MJ, Earles JA, Teachout MS. Predicting job performance: Not much more than g. *Journal of applied psychology.* 1994;79(4):518.
5. Crumley LM, Pierce LG, Schwalm RC, Coke JS, Brown JC. *Predicting target detection performance using the armed services vocational aptitude battery subtests and cognitive factor tests.* ARMY RESEARCH INST FOR THE BEHAVIORAL AND SOCIAL SCIENCES ALEXANDRIA VA;1992.
6. Parasuraman R. Assaying individual differences in cognition with molecular genetics: theory and application. *Theoretical Issues in Ergonomics Science.* 2009;10(5):399-416.
7. Figueroa IJ, Youmans RJ. Individual differences in cognitive flexibility predict performance in vigilance tasks. Paper presented at: Proceedings of the human factors and ergonomics society annual meeting2012.
8. Finomore V, Matthews G, Shaw T, Warm J. Predicting vigilance: a fresh look at an old problem. *Ergonomics.* 2009;52(7):791-808.
9. Schwark J, Sandry J, Dolgov I. Evidence for a positive relationship between working-memory capacity and detection of low-prevalence targets in visual search. *Perception.* 2013;42(1):112-114.
10. Peltier C, Becker MW. Working Memory Capacity Predicts Selection and Identification Errors in Visual Search. *Perception.* 2017;46(1):109-115.
11. Hattenschwiler N, Merks S, Sterchi Y, Schwaninger A. Traditional Visual Search vs. X-Ray Image Inspection in Students and Professionals: Are the Same Visual-Cognitive Abilities Needed? *Front Psychol.* 2019;10:525.
12. Biggs AT, Clark K, Mitroff SR. Who should be searching? Differences in personality can affect visual search accuracy. *Personality and Individual Differences.* 2017;116:353-358.
13. Biggs AT, Cain MS, Clark K, Darling EF, Mitroff SR. Assessing visual search performance differences between Transportation Security Administration Officers and nonprofessional visual searchers. *Visual Cognition.* 2013;21(3):330-352.
14. Santiago-Espada Y, Myer RR, Latorella KA, Comstock Jr JR. The multi-attribute task battery ii (matb-ii) software for human performance and workload research: A user's guide. 2011.
15. Singh AL, Tiwari T, Singh IL. Performance feedback, mental workload and monitoring efficiency. *J Indian Acad Appl Psychol.* 2010;36(1):151-158.
16. Morgan B, D'Mello S, Abbott R, Radvansky G, Haass M, Tamplin A. Individual differences in multitasking ability and adaptability. *Hum Factors.* 2013;55(4):776-788.
17. Basner M, Savitt A, Moore TM, et al. Development and Validation of the Cognition Test Battery for Spaceflight. *Aerosp Med Hum Perform.* 2015;86(11):942-952.

18. Gur RC, Ragland JD, Moberg PJ, et al. Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacology.* 2001;25(5):766-776.
19. Glahn DC, Gur RC, Ragland JD, Censits DM, Gur RE. Reliability, performance characteristics, construct validity, and an initial clinical application of a visual object learning test (VOLT). *Neuropsychology.* 1997;11(4):602-612.
20. Ragland JD, Turetsky BI, Gur RC, et al. Working memory for complex figures: an fMRI comparison of letter and fractal n-back tasks. *Neuropsychology.* 2002;16(3):370-379.
21. Glahn DC, Cannon TD, Gur RE, Ragland JD, Gur RC. Working memory constrains abstraction in schizophrenia. *Biol Psychiatry.* 2000;47(1):34-42.
22. Moore TM, Scott JC, Reise SP, et al. Development of an abbreviated form of the Penn Line Orientation Test using large samples and computerized adaptive test simulation. *Psychol Assess.* 2015;27(3):955-964.
23. Benton AL, Varney NR, Hamsher KD. Visuospatial judgment. A clinical test. *Arch Neurol.* 1978;35(6):364-367.
24. Gur RC, Sara R, Hagendoorn M, et al. A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies. *J Neurosci Methods.* 2002;115(2):137-143.
25. Gur RC, Richard J, Hughett P, et al. A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: standardization and initial construct validation. *J Neurosci Methods.* 2010;187(2):254-262.
26. Usui N, Haji T, Maruyama M, et al. Cortical areas related to performance of WAIS Digit Symbol Test: a functional imaging study. *Neurosci Lett.* 2009;463(1):1-5.
27. Lejuez CW, Read JP, Kahler CW, et al. Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART). *J Exp Psychol Appl.* 2002;8(2):75-84.
28. Drummond SP, Bischoff-Grethe A, Dinges DF, Ayalon L, Mednick SC, Meloy MJ. The neural basis of the psychomotor vigilance task. *Sleep.* 2005;28(9):1059-1068.
29. Peltier C, Becker MW. Individual differences predict low prevalence visual search performance. *Cogn Res Princ Implic.* 2017;2(1):5.
30. Manly T, Robertson IH, Galloway M, Hawkins K. The absent mind: further investigations of sustained attention to response. *Neuropsychologia.* 1999;37(6):661-670.
31. Adamo SH, Cain MS, Mitroff SR. An individual differences approach to multiple-target visual search errors: How search errors relate to different characteristics of attention. *Vision Res.* 2017;141:258-265.
32. Peltier C, Becker MW. Decision processes in visual search as a function of target prevalence. *J Exp Psychol Hum Percept Perform.* 2016;42(9):1466-1476.
33. Basner M, Moore TM, Hermosillo E, et al. Cognition Test Battery Performance Is Associated with Simulated 6df Spacecraft Docking Performance. *Aerosp Med Hum Perform.* 2020;91(11):861-867.
34. Tottenham N, Tanaka JW, Leon AC, et al. The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Res.* 2009;168(3):242-249.
35. Hambrick DZ, Altmann EM. The role of placekeeping ability in fluid intelligence. *Psychon Bull Rev.* 2015;22(4):1104-1110.
36. Burgoyne AP, Hambrick DZ, Altmann EM. Placekeeping ability as a component of fluid intelligence: Not just working memory capacity. *The American Journal of Psychology.* 2019;132(4):439-449.

37. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: *Advances in psychology.* Vol 52. Elsevier; 1988:139-183.
38. Bernhardt KA, Salomon KA, Ferraro FR, et al. Individual differences in dynamic multitasking performance. Paper presented at: Proceedings of the Human Factors and Ergonomics Society Annual Meeting2016.
39. Kennedy L, Parker SH. Making MATB-II medical: pilot testing results to determine a novel lab-based, stress-inducing task. Paper presented at: Proceedings of the international symposium on human factors and ergonomics in health care2017.
40. Gur RC, Richard J, Calkins ME, et al. Age group and sex differences in performance on a computerized neurocognitive battery in children age 8-21. *Neuropsychology.* 2012;26(2):251-265.
41. Moore TM, Basner M, Nasrini J, et al. Validation of the Cognition Test Battery for Spaceflight in a Sample of Highly Educated Adults. *Aerosp Med Hum Perform.* 2017;88(10):937-946.
42. Voyer D, Voyer S, Bryden MP. Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychol Bull.* 1995;117(2):250-270.
43. Williams LM, Mathersul D, Palmer DM, Gur RC, Gur RE, Gordon E. Explicit identification and implicit recognition of facial emotions: I. Age effects in males and females across 10 decades. *J Clin Exp Neuropsychol.* 2009;31(3):257-277.
44. Jewett ME, Dijk DJ, Kronauer RE, Dinges DF. Dose-response relationship between sleep duration and human psychomotor vigilance and subjective alertness. *Sleep.* 1999;22(2):171-179.
45. Oberlander EM, Oswald FL, Hambrick DZ, Jones LA. *Individual difference variables as predictors of error during multitasking.* NAVY PERSONNEL RESEARCH STUDIES AND TECHNOLOGY MILLINGTON TN;2007.
46. Hambrick DZ, Oswald FL, Darowski ES, Rench TA, Brou R. Predictors of multitasking performance in a synthetic work paradigm. *Applied cognitive psychology.* 2010;24(8):1149-1167.
47. Wechsler D. The measurement of adult intelligence. 1944.
48. Roberts RD, Goff GN, Anjoul F, Kyllonen PC, Pallier G, Stankov L. The armed services vocational aptitude battery (ASVAB): Little more than acculturated learning (Gc)!? *Learning and Individual Differences.* 2000;12(1):81-103.
49. Hambrick DZ, Rench TA, Poposki EM, et al. The relationship between the ASVAB and multitasking in Navy sailors: A process-specific approach. *Military Psychology.* 2011;23(4):365-380.
50. Giwa A, Adeyi AA, Giwa SO. Empirical modelling and optimization of PAME reactive distillation process using Minitab. *International Journal of Scientific & Engineering Research.* 2015;6(6):1-12.
51. Grekousis G. *Spatial analysis methods and practice: describe–explore–explain through GIS.* Cambridge University Press; 2020.
52. Drew T, Williams LH. Simple eye-movement feedback during visual search is not helpful. *Cogn Res Princ Implic.* 2017;2(1):44.
53. Peltier C, Becker MW. Eye movement feedback fails to improve visual search performance. *Cogn Res Princ Implic.* 2017;2(1):47.

54. Wolfe JM, Horowitz TS, Van Wert MJ, Kenner NM, Place SS, Kibbi N. Low target prevalence is a stubborn source of errors in visual search tasks. *J Exp Psychol Gen.* 2007;136(4):623-638.
55. Schwark J, Sandry J, Macdonald J, Dolgov I. False feedback increases detection of low-prevalence targets in visual search. *Atten Percept Psychophys.* 2012;74(8):1583-1589.
56. Taylor JET, Hilchey MD, Weidler BJ, Pratt J. Eliminating the Low Prevalence Effect in Visual Search with a Remarkably Simple Strategy. 2021.
57. Ahonniska J, Ahonen T, Aro T, Tolvanen A, Lyytinen H. Repeated assessment of the Tower of Hanoi test: reliability and age effects. *Assessment.* 2000;7(3):297-310.
58. Hambrick DZ, Altmann EM, Oswald FL, Meinz EJ, Gobet F. Facing facts about deliberate practice. *Front Psychol.* 2014;5:751.
59. Meinz EJ, Hambrick DZ. Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: the role of working memory capacity. *Psychol Sci.* 2010;21(7):914-919.
60. Kyllonen PC, Kell H. Ability Tests Measure Personality, Personality Tests Measure Ability: Disentangling Construct and Method in Evaluating the Relationship between Personality and Ability. *J Intell.* 2018;6(3).

Appendix

Supplemental Table 1. MATB-II Mean and Standard Deviation scores (SD in parentheses).

	Block I 0-09 min	Block II 10-19 min	Block III 20-29 min	Block IV 30-39 min	Block V 40-49 min	Block VI 50-59 min	Overall
Raw Scores							
SYSMON-Acc	0.93 (0.13)	0.97 (0.06)	0.96 (0.06)	0.97 (0.05)	0.98 (0.04)	0.97 (0.06)	0.96 (0.32)
SYSMON-RT	2.98 (1.06)	2.92 (0.70)	2.96 (0.88)	2.96 (0.87)	2.88 (0.87)	2.84 (0.88)	2.89 (0.47)
TRACK	27.60 (5.382)	28.56 (7.42)	27.61 (7.06)	28.61 (6.60)	28.08 (6.49)	28.46 (7.08)	28.18 (6.37)
COMMS-Acc	0.95 (0.08)	0.98 (0.05)	0.95 (0.11)	0.94 (0.10)	0.96 (0.06)	0.98 (0.05)	0.96 (0.05)
COMMS-RT	2.26 (0.83)	2.33 (0.85)	2.06 (0.70)	2.06 (0.84)	1.80 (0.81)	1.56 (0.57)	2.03 (0.48)
RESMAN	-213.19 (560.52)	-797.81 (989.91)	-907.01 (861.31)	-845.92 (947.68)	-1169.54 (667.05)	-1305.84 (777.72)	-821.12 (576.44)
Z-Scores							
SYSMON-Acc	0.20 (1.95)	0.61 (0.84)	0.53 (0.91)	0.68 (0.78)	0.83 (0.67)	0.65 (0.93)	0.57 (0.48)
SYSMON-RT	4.17 (3.18)	4.02 (2.10)	4.13 (2.63)	4.13 (2.62)	3.89 (2.61)	3.77 (2.63)	3.92 (1.39)
TRACK	-0.22 (0.70)	-0.11 (0.89)	-0.22 (0.85)	-0.10 (0.79)	-0.16 (0.78)	-0.12 (0.85)	-0.15 (0.79)
COMMS-Acc	-0.64 (1.76)	0.08 (1.17)	-0.53 (2.54)	-0.73 (2.36)	-0.23 (1.42)	0.19 (1.05)	-0.37 (1.02)
COMMS-RT	-0.45 (0.55)	-0.40 (0.56)	-0.58 (0.46)	-0.58 (0.55)	-0.74 (0.53)	-0.91 (0.38)	-0.60 (0.31)
RESMAN	-0.54 (1.55)	-2.64 (4.00)	-2.87 (3.59)	-3.07 (3.44)	-3.98 (2.61)	-4.51 (3.39)	-2.19 (2.55)
Composite	-1.30 (3.52)	**-3.12 (4.70)***	**-3.76 (4.48)***	**-3.65 (4.54)***	**-4.34 (3.52)***	**-4.68 (3.68)***	-3.26 (3.31)

Reaction time (RT) and root mean square error (RMSE) z-scores are shown not inverted. The top half of the table shows the mean and standard deviation (SD) of the raw scores in each MATB-II sub-task, of each time block (a ten-minute segment). The bottom half of the table shows the mean performance (as weighted, normalized z-scores) of each MATB-II sub-task, of each block. The referent scores from which the z-score is derived are the Day 1 scores. Composite z-scores denoted in bold font and with an asterisk (*) were significantly different from Block I ($p < .05$)

Supplemental Table 2. Cognitive Test Battery: Mean and Standard Deviation scores (SD in parentheses) for the 20 predictor sub-tasks

Dependent Measure	Mean (SD)
MPT: reaction time (ms)	772.41 (73.02)
VOLT: accuracy (proportion correct)	0.82 (0.09)
VOLT: reaction time (ms)	3401.40 (885.18)
F2B: accuracy (proportion hits)	0.71 (0.12)
F2B: incorrect (proportion misses)	0.27 (0.11)
F2B: reaction time (ms)	655.65 (98.56)
WCS: accuracy (proportion correct)	0.81 (0.12)
WCS: reaction time (ms)	1458.67 (269.46)
WCS: number of perseverative error trials	1.00 (0.72)
LOT: error (degrees)	1.29 (0.49)
LOT: number of clicks	6.99 (1.03)
ERT: accuracy (proportion correct)	0.90 (0.05)
ERT: reaction time (ms)	2254.80 (504.25)
MRT: accuracy (proportion correct)	0.25 (0.03)
MRT: reaction time (ms)	21269.52 (8790.82)
DSST: number of correct trials	48.25 (7.72)
DSST: reaction time (ms)	1831.06 (287.81)
BART: average number of pumps	32.61 (10.61)
PVT: reaction time (ms)	284.68 (45.66)
PVT: number of lapses	1.75 (3.22)

Supplemental Table 3. MATB-II composite performance (total) best subset regression results for models ranging from 2 to 10 maximum predictor variables

Max	Predictors				Overall Model			Cross-Validation		
Number of Predictors	Predictors	Estimate (B)	Std. Error	p-value	R^2	Adj. R^2	p-value	pred R^2	Adj R^2 / R^2	MSE
2	F2B reaction time	-0.012	0.006	.05	0.43	0.38	.003	0.27	0.87	10
	MPT reaction time	-0.02	0.008	**.02***	Significant predictors: 1/2 = .50					
3	ERT reaction time	0.001	0.001	.25	0.47	0.39	.005	0.67	0.82	20.69
	F2B reaction time	-0.013	0.006	**.04***	Significant predictors: 2/3 = .67					
	MPT reaction time	-0.025	0.009	**.01***						
4	DDST # of correct trials	-0.151	0.102	.16	0.52	0.42	.006	0.21	0.8	25.6
	F2B reaction time	-0.017	0.006	**.01***	Significant predictors: 2/4 = .50					
	MPT reaction time	-0.031	0.011	**.01***						
	LOT avg. error	1.744	1.208	.165						
5	DDST # of correct trials	-0.178	0.102	.10	0.57	0.45	.006	0.16	0.78	30.9
	F2B reaction time	-0.014	0.007	**.05***	Significant predictors: 2/5 = .40					
	MPT reaction time	-0.033	0.01	**.01***						
	LOT avg. error	1.682	1.179	.17						
	PVT reaction time	-0.018	0.013	.18						
6	DDST # of correct trials	-0.142	0.104	.19	0.6	0.6	.008	0.04	0.76	33.53
	F2B reaction time	-0.015	0.006	**.04***	Significant predictors: 2/6 = .33					
	MPT reaction time	-0.027	0.011	**.03***						
	LOT avg. error	2.017	1.192	.11						
	PVT reaction time	-0.024	0.013	.09						
	MRT accuracy	-26.285	21.08	.23						
7	WSC reaction time	0.004	0.003	.16	0.64	0.48	.01	0.11	0.74	39.01
	BART avg. # of pumps	-0.116	0.063	.08	Significant predictors: 2/7 = .29					
	ERT reaction time	0.002	0.001	.22						
	F2B reaction time	-0.015	0.006	**.03***						
	LOT avg. error	1.754	1.15	.15						
	MPT reaction time	-0.034	0.01	**.002***						
	PVT reaction time	-0.019	0.012	.14						
8	BART avg. # of pumps	-0.08	0.053	.15	0.68	0.51	.01	0.01	0.74	48.78
	ERT accuracy	-23.89	12.51	.08	Significant predictors: 6/8 = **.75**					
	F2B Hits	-33.118	14.54	**.04***						
	F2B Misses	-45.827	16.56	**.01***						
	F2B reaction time	-0.019	0.006	**.01***						
	LOT avg. error	3.048	1.274	**.03***						
	MRT accuracy	-59.588	19.07	**.01***						
	PVT reaction time	-0.04	0.014	**.01***						
9	BART avg. # of pumps	-0.087	0.053	.12	0.71	0.52	.013	-0.1	0.72	62.18
	ERT accuracy	-22.789	12.38	.09	Significant predictors: 5/9 = .56					
	F2B Hits	-29.223	14.73	.07						
	F2B Misses	-38.453	17.51	**.05***						
	F2B reaction time	-0.017	0.007	**.02***						
	LOT avg. error	3.021	1.258	**.03***						
	MPT reaction time	-0.01	0.009	.26						
	MRT accuracy	-48.887	20.91	**.04***						
	PVT reaction time	-0.038	0.014	**.02***						
10	DDST # of correct trials	-0.12	0.105	.27	0.74	0.53	.017	-0.2	0.71	76.53

BART avg. # of pumps	-0.079	0.053	.16	Significant predictors: 3/19 = .30
ERT accuracy	-21.256	12.32	.11	
F2B Hits	-25.214	14.98	.12	
F2B Misses	-35.624	17.49	.06	
F2B reaction time	-0.019	0.007	**.02***	
LOT avg. error	3.067	1.245	**.03***	
MPT reaction time	-0.019	0.011	.12	
MRT accuracy	-41.94	21.55	.07	
PVT reaction time	-0.037	0.014	**.02***	

Significant predictors (p<.05) are indicated in bold with an asterisk (*)

Supplemental Table 4. MATB-II sustained attention performance (slope) best subset regression results for models ranging from 2 to 10 maximum predictor variables

Max Number of Predictors	Predictors				Overall Model			Cross-Validation		
	Predictors	Estimate (B)	Std. Error	p-value	R^2	Adj. R^2	p-value	pred R^2	Adj R^2 / R^2	MSE
2	WSC reaction time	<0.01	<0.01	.16	0.27	0.20	.039	-0.01	0.72	0.97
	BART avg. # of pumps	0.041	0.015	**.01***	Significant predictors: 1/2 = .5					
3	WSC reaction time	-0.002	0.001	**.02***	0.40	0.31	.015	0.1	0.76	1.25
	BART avg. # of pumps	0.062	0.017	**.002***	Significant predictors: 3/3 = **1.00**					
	DDST # of correct trials	-0.053	0.025	**.05***						
4	WSC reaction time	-0.002	0.001	**.01***	0.49	0.38	.01	0.17	0.77	1.40
	BART avg. # of pumps	0.069	0.017	**.001***	Significant predictors: 3/4 = .75					
	DDST # of correct trials	-0.075	0.027	**.01***						
	F2B Misses	-2.473	1.355	.08						
5	WSC reaction time	-0.002	0.001	**.01***	0.54	0.41	.011	0.13	0.75	1.71
	BART avg. # of pumps	0.069	0.016	**<.001***	Significant predictors: 3/4 = .60					
	DDST # of correct trials	-0.074	0.026	**.01***						
	F2B Misses	-2.117	1.351	.13						
	LOT avg. error	0.365	0.267	.19						
6	WSC reaction time	-0.002	0.001	**.003***	0.61	0.47	.007	0.07	0.76	2.04
	BART avg. # of pumps	0.061	0.016	**.001***	Significant predictors: 3/6 = .50					
	DDST # of correct trials	-0.072	0.025	**.01***						
	F2B Misses	-2.403	1.256	.07						
	MRT accuracy	10.384	4.93	.05						
	VOLT accuracy	-3.087	1.605	.07						
7	WSC reaction time	-0.002	0.001	**.002***	0.67	0.53	.005	0.12	0.78	2.39
	BART avg. # of pumps	0.065	0.015	**<.001***	Significant predictors: 4/7 = .57					
	DDST # of correct trials	-0.065	0.024	**.01***						
	F2B Misses	-2.27	1.193	.075						
	MRT accuracy	12.482	4.83	**.02***						
	VOLT accuracy	-2.958	1.522	.07						
	PVT reaction time	0.005	0.003	.11						
8	WSC perseverate errors	0.422	0.277	.15	0.74	0.61	.323	0.11	0.8	2.25
	ERT accuracy	11.02	5.157	**.05***	Significant predictors: 2/8 = .25					
	ERT reaction time	0.001	0.001	**.05***						
	F2B Hits	-1.866	1.564	.25						
	MPT reaction time	-0.007	0.004	.07						
	MRT accuracy	12.17	6.48	.08						
	PVT reaction time	0.02	0.011	.09						

	PVT # of lapses	-0.165	0.145	.27							
9	WSC perseverate errors	0.492	0.168	**.01***	0.80	0.67	.001	0.38	0.83	2.44
	ERT accuracy	18.38	3.431	**<.001***	Significant predictors: 9/9 = 1.00					
	ERT reaction time	0.002	0.0003	**<.001***						
	F2B Hits	-3.411	0.993	**.004***						
	MPT reaction time	-0.013	0.002	**<.001***						
	MRT accuracy	25.95	4.739	**<.001***						
	PVT reaction time	0.032	0.007	**<.001***						
	PVT # of lapses	-0.32	0.092	**.004***						
	VOLT accuracy	-7.272	1.402	**<.001***						
10	WSC perseverate errors	0.532	0.143	**.003***	0.86	0.76	.0004	0.56	0.87	1.70
	ERT accuracy	17.38	0.0003	**<.001***	Significant predictors: 10/10 = 1.00					
	ERT reaction time	0.002	0.0003	**<.001***						
	F2B Hits	-9.386	2.495	**.002***						
	F2B Misses	-7.554	2.969	**.02***						
	MPT reaction time	-0.011	0.002	**<.001***						
	MRT accuracy	23.22	4.16	**<.001***						
	PVT reaction time	0.033	0.006	**<.001***						
	PVT # of lapses	-0.355	0.08	**.001***						
	VOLT accuracy	-7.86	1.211	**<.001***						

Significant predictors (p<.05) are indicated in bold with an asterisk (*)