Published in final edited form as:

Aerosp Med Hum Perform. 2020 November 01; 91(11): 861–867. doi:10.3357/AMHP.5602.2020.

Cognition Test Battery Performance is Associated with Simulated 6 DF Spacecraft Docking Performance

Mathias Basner, MD, PhD, MSc¹, Tyler M. Moore, PhD², Emanuel Hermosillo¹, Jad Nasrini¹, David F. Dinges, PhD¹, Ruben C. Gur, PhD², Bernd Johannes, PhD³

¹Division of Sleep and Chronobiology, Department of Psychiatry, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA

²Brain Behavior Laboratory, Department of Psychiatry, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA

³Institute of Aerospace Medicine, German Aerospace Center (DLR), Cologne, Germany

Abstract

Introduction: Environmental and operational stressors commonly encountered in spaceflight can affect astronaut cognitive performance. It is currently unclear how performance decrements on test batteries that assess individual cognitive domains translate to complex operational performance.

Methods: N=30 healthy adults (mean±SD age 33.5±7.1 years, range 25–48 years; 16 male) with demographic characteristics similar to astronauts performed all 10 tests of the Cognition test battery as well as a simulated 6df spacecraft docking task 15 times. Performance on 60 Cognition outcome variables was rank-correlated with 6df docking performance individually as well as in models containing up to 12 predictors after accounting for sex, age, and study design effects.

Results: Average response time on the Digit Symbol Substitution Test (DSST) – a measure of processing speed requiring complex scanning, visual tracking and working memory – was the best individual predictor of 6df docking performance (unadjusted $r=-0.550$; semi-partial cross-validated $R^2=0.244$). Furthermore, higher levels of spatial orientation efficiency and vigilant attention, lower levels of impulsivity, and faster response speed were associated with higher 6df performance, while sensorimotor speed, memory and risk decision making were less relevant. After semi-partial cross-validation, a model with three Cognition outcomes (DSST average response time, Abstract Matching accuracy, and conservative response bias on the Fractal 2-Back test) explained 30% of the variance in 6df performance.

Conclusions: This study demonstrates direct links between performance on tests designed to assess specific cognitive domains and complex operational docking performance.

Keywords

cognitive test; spaceflight; astronaut; microgravity

Introduction

Spacecraft are considered isolated, confined, and extreme (ICE) environments. As small operational errors can cause catastrophic events, astronauts are required to maintain high levels of cognitive performance at all times. Astronaut Scott Kelly noted the following in his book about his 340-day International Space Station (ISS) mission:

“Perhaps the most dangerous symptom is impairment to cognitive function – we have to be able to perform tasks that require a high degree of concentration and attention to detail at a moment’s notice, and in an emergency, which can happen anytime, we need to be able to do those tasks right at the first time. Losing just a fraction of our ability to focus, make calculations, or solve problems could cost our lives.”²²

In addition, the spacecraft environment contains several environmental (e.g., microgravity, radiation, confinement), psychological (e.g., isolation, threat to life, communication delays) and operational (e.g., spacecraft dockings, extravehicular activities) stressors that have the potential to compromise astronaut cognitive functions and performance.²⁸ It is therefore important to objectively monitor astronaut cognitive performance on a regular basis during long-duration space missions.

For this purpose, NASA has been using the WinSCAT battery, which contains a subset of six Automated Neuropsychological Assessment Metrics (ANAM) tests, for over 20 years to assess astronaut cognitive performance.²¹ While these tests are well validated, they have two disadvantages: (1) the ANAM tests selected for WinSCAT assess a relatively narrow spectrum of cognitive domains, heavily focusing on working memory; and (2) several tests are too easy for the high-performing astronaut population, introducing ceiling effects and potentially reducing astronaut motivation. For these reasons, we recently developed a new cognitive test battery specifically designed for the high-performing astronaut population. This battery - called *Cognition* contains 10 brief tests that cover a range of cognitive domains relevant for mission success and is based on tests known to engage specific brain systems during functional neuroimaging (see Methods section for a description of the 10 *Cognition* tests)^{4, 16, 27}. We generated 15 unique versions of *Cognition* that allow for repeated testing in astronauts.

The main advantage of administering a number of performance tests that probe a range of cognitive abilities is that the pattern of performance decrements across tests allows for the localization of brain systems primarily affected and may guide the process of identifying stressors that affect the same brain systems in controlled experiments on the ground. A potential disadvantage is that it is unclear how performance decrements on a single cognitive test or multiple tests translate to operational performance.

Studies investigating the relationship between performance on cognitive tests with operational outcomes are scarce, especially for spaceflight. We therefore conducted a study in which subjects repeatedly performed the *Cognition* test battery as well as a complex 6 degrees-of-freedom spacecraft docking task (6df) that simulates docking operations on the ISS and is used by cosmonauts on the ISS to stay proficient on the task¹⁸. The main

objective of the study was to investigate which subset of *Cognition* test outcomes best explained performance on the operational task, and how much of the variance in operational performance was accounted for by *Cognition*.

Methods

Subjects

A total of N=30 healthy adult subjects (mean±SD age 33.5±7.1 years, range 25–48 years; 16 male) with demographic characteristics similar to astronauts were recruited for the study through on-campus and online postings; all 30 completed the study. Subjects filled out a screening questionnaire prior to enrollment. They had to be proficient in English, between 25 and 55 years old, and have at least a Master's degree (or equivalent). Subjects with a history of a medical disorder that may significantly impact brain function (e.g., neurological disorders, severe head trauma) as well as subjects with a diagnosed psychiatric illness were excluded from study participation. The study was reviewed by the Institutional Review Board of the University of Pennsylvania and considered exempt. Written informed consent was obtained from study participants prior to data collection. Subjects were compensated for each session they attended. Depending on group assignment they received up to \$390 if they completed all test sessions.

Materials and Equipment

The Cognition test battery was administered on a calibrated laptop computer (Dell Latitude E6430 with a 14" screen diagonal, 16:9 aspect ratio). A detailed description of the 10 *Cognition* tests can be found in Basner et al.⁴. Screenshots of the 10 tests can be found in supplementary Figure S1. Here, we provide a brief characterization of each test (modified from Basner et al.⁵). The 10 tests were always performed in the order they are listed below.

The **Motor Praxis Task (MP)**¹⁵ was administered at the start of testing to ensure that participants have sufficient command of the computer interface, and immediately thereafter as a measure of sensorimotor speed. Participants were instructed to click on squares that appear in random locations on the screen, with each successive square smaller and thus more difficult to track. Performance was assessed by the speed with which participants click each square.

The **Visual Object Learning Test (VOLT)**¹³ assessed participant memory for complex figures. Participants were asked to memorize 10 sequentially displayed three-dimensional shapes. Later, they were instructed to select the objects they memorized from a set of 20 such objects, also sequentially presented, half of which were from the learning set and half new.

The **Fractal 2-Back (F2B)**²⁵ is a nonverbal variant of the Letter 2-Back. N-back tasks have become standard probes of the working memory system and activate canonical working memory brain areas. The F2B consisted of the sequential presentation of a set of figures (fractals), each potentially repeated multiple times. Participants were instructed to respond when the current stimulus matched the stimulus displayed two figures ago. The current implementation used 62 consecutive stimuli.

The **Abstract Matching (AM)** test¹⁴ is a validated measure of the abstraction and flexibility components of executive function, including an ability to discern general rules from specific instances. The test paradigm presented subjects with two pairs of objects at the bottom left and right of the screen, varied on perceptual dimensions (e.g., color and shape). Subjects were presented with a target object in the upper middle of the screen that they had to classify as belonging more with one of the two pairs, based on a set of implicit, abstract rules. The current implementation used 30 consecutive stimuli.

The **Line Orientation Test (LOT)**⁹ is a measure of spatial orientation and derived from the well-validated Judgment of Line Orientation Test. The LOT format consisted of presenting two lines at a time, one stationary while the other could be rotated by clicking an arrow. Participants could rotate the movable line until they perceived it to be parallel to the stationary line. The implementation used in this study had 12 consecutive line pairs that varied in length and orientation.

The **Emotion Recognition Task (ERT)**¹⁶ is a measure of emotion recognition. It presented subjects with photographs of professional actors (adults of varying age and ethnicity) portraying emotional facial expressions of varying types and intensities (biased towards lower intensities, and with the prevalence of emotion categories balanced within each version of the test). Subjects were given a set of emotion labels (“happy”; “sad”; “angry”; “fearful”; and “no emotion”) and had to select the label that correctly described the expressed emotion. The implementation for the study used 40 consecutive stimuli, with 8 stimuli each representing one of the above 5 emotion categories.

The **Matrix Reasoning Test (MRT)**¹⁵ is a measure of abstract reasoning and consists of increasingly difficult pattern matching tasks. It is analogous to Raven’s Progressive Matrices,²⁶ a well-known measure of general intelligence. The test consisted of a series of patterns, overlaid on a grid. One element from the grid was missing and the participant had to select the element that fit the pattern from a set of several options. The implementation in the study used 12 consecutive stimuli.

The **Digit-Symbol Substitution Task (DSST)**²⁹ is a measure of processing speed requiring complex scanning, visual tracking and working memory, and a computerized adaptation of a paradigm used in the Wechsler Adult Intelligence Scale (WAIS-III). The DSST required the participant to refer to a displayed legend relating each of the digits one through nine to specific symbols. One of the nine symbols appeared on the screen and the participant had to type the corresponding number using the keyboard as quickly as possible. The test duration was fixed at 90 s, and the legend key was randomly re-assigned with each administration.

The **Balloon Analog Risk Test (BART)** is a validated assessment of risk taking behavior.²³ The BART required participants to either inflate an animated balloon or stop inflating and collect a reward. Participants were rewarded in proportion to the final size of each balloon, but a balloon popped after a hidden number of pumps, which changed across stimuli.²³ The implementation in the study used 30 consecutive stimuli. The average tendency of balloons to pop was systematically varied between test administrations. This variation required subjects to adjust the level of risk based on the behavior of the balloons.

The **Psychomotor Vigilance Test (PVT)** is a validated measure of sustained attention based on reaction time (RT) to visual stimuli that occur at random inter-stimulus intervals.^{2, 11} Subjects were instructed to monitor a box on the screen and press the space bar when an incrementing millisecond counter appeared in the box. The reaction time was then displayed for one second. Subjects were instructed to be as fast as possible without hitting the spacebar in the absence of a stimulus (i.e., false starts or errors of commission). The PVT is a sensitive measure of vigilant attention, and has been well-established as a tool to detect acute and chronic sleep deprivation and circadian misalignment, conditions highly prevalent in spaceflight.¹ The PVT has negligible aptitude and learning effects,² and is ecologically relevant as sustained attention deficits and slow reactions affect many real-world tasks (e.g., operating a moving vehicle).¹⁰ In the current study, *Cognition* contained a validated 3-min. brief PVT-B with 2–5 s inter-stimulus intervals and a 355 ms lapse threshold.³ For the analyses described below, we used 60 *Cognition* outcome variables, including one standard speed and one standard accuracy outcome for each test. These variables are described in detail in the Supplement (Table S1). We first calculated average performance scores across the 15 administrations of *Cognition* for each subject and each outcome variable. Subjects were then ranked according to their average performance score in ascending order (i.e., the lowest score was ranked 1 and the highest score was ranked 30). The average rank was assigned to tied scores.

After each *Cognition* test battery, a Simulated 6 Degrees of Freedom Spacecraft Docking Task (6df) training program was administered to teach participants the use of the hand controllers and to guide them through tasks with ever increasing difficulty.^{20, 18} The training program is described in detail in Johannes et al.¹⁹ and screenshots are provided in supplementary Figure S2. The 6df task simulates the manual control of an object in space with six degrees of freedom. A chain of successive semi-transparent ellipsoid rings displays the required path towards the target object. The three dimensional path of ellipsoids has to be traversed while looking from inside the moving object to outside on a camera monitor. The 6df course started with familiarizing the subject with the different functionalities of the two hand controls for translation and orientation. Basic elements of the control were practiced separately: moving along the object axes, oblique to the moved object axes, curves, final approach, and contact (capture, docking). The degrees of freedom that had to be controlled were increased systematically. Whether a participant progressed to the next level was decided by a rule system based on the performance of the subject.¹⁹ A task level was considered complete if the subject docked with an accuracy of at least 95%. The training program terminated automatically after 15 minutes. At the beginning of the next session, the training program first repeated the last training session successfully completed by the subject before administering the next task level. Ranking of 6df performance was based on the highest task level completed (the highest task level completed in this study was 40 out of 62). In subjects who completed the same task level, rank was determined by average docking accuracy (without including the last flight of each day).

Procedure

Participants performed the *Cognition* test battery⁴ and 6df a total of 15 times in an office located in the Unit for Experimental Psychiatry at the University of Pennsylvania School of

Medicine. The main goal of the study was to disentangle practice and stimulus set difficulty effects of the *Cognition* battery.^{6, 7} Before the first test administration, subjects watched a standardized familiarization video and were asked to perform a brief practice version of 8 out of the 10 *Cognition* tests before each of those tests for the first administration (the VOLT and BART have no practice versions). The practice versions were available before their respective tasks throughout the study, but were not required starting with the second battery administration. In the video, the principal investigator performed a short version of all 10 *Cognition* tests after reading out loud the standardized instructions for each test. During all test administrations, subjects were supervised by a research coordinator who could answer questions and address technical problems at any time. Between test days, subjects went about their normal activities. They were not instructed to follow a specific sleep-wake schedule, and they were free to choose the time of day for testing (i.e., they did not perform *Cognition* at the same time of day every time they were tested). However, 85.4% of testing sessions started between 9 am and 4 pm (earliest 7:10 am, latest 5:59 pm).

The 15 versions of *Cognition* were administered in a randomized but balanced fashion, i.e., each subject performed each version of *Cognition* exactly once, and each version of *Cognition* was administered in each administration order position exactly once. A secondary objective of the study was to test the effect of the duration of the interval between two *Cognition* administrations on practice effects. Subjects performed *Cognition* either with 10 days between test administrations (long group; N=14 subjects completed the protocol) or 5 days between test administrations (short group; N=16 subjects completed the protocol). In the short group, there had to be at least one day without testing between study visits.

Statistical Analysis

It was determined whether 6df performance rank differed between participant sex and administration interval groups (Welch's t-test in R), and whether age and 6df performance rank correlated. Bivariate correlations and semi-partial bivariate correlations (given age, sex, and administration interval group) between 6df performance rank and the ranked 60 *Cognition* outcomes were calculated. Finally, cross-validated best-subset regression¹² was used to build a model predicting 6df performance using only the most powerful predictors. In best subset regression, a number of variables is specified by the user (e.g. 2 variables) and the algorithm tests every possible combination of that many variables, checking the R^2 for each combination to evaluate the model quality. For example, if we had 20 variables total and we specified that we wanted 2 variables in the final model, the best-subset regression algorithm would test $2^{20} = 1,048,576$ models. In our analyses, we performed best subset regression for models containing between 2 (high parsimony) and 12 (low parsimony) variables. With these optimal models built, the predictive utility of the models was next tested using 10-fold cross-validation—i.e. estimated model in 90% of sample and tested in 10%, and repeated that procedure 10 times such that each person received a predicted value from a model estimated without him/her. The cross-validated model with the highest R-squared value was selected as the optimal model. Statistical significance was assumed at $p < 0.05$. All analyses were performed in R version 3.6.1. Study raw data can be requested through NASA's Life Science Data Archive (LSDA).

Results

Male subjects (average rank 20.6) performed significantly better on 6df than females subjects (average rank 9.6; $p=0.0002$, Welch two sample t-test), while those with long administration intervals (average rank 15.8) did not differ in 6df performance from those with short administration intervals (average rank 15.3; $p=0.8714$, Welch two sample t-test). Age showed a trend to negatively correlate with 6df performance (Pearson's $\rho=-0.221$; $p=0.2408$).

The rank for Cognition speed across cognitive domains decreased significantly with age (0.33 ranks per year; $p=0.0181$), while neither sex ($p=0.6922$) nor administration interval ($p=0.1285$) was significantly associated with Cognition speed. Age, sex and administration interval were not associated with the rank of Cognition accuracy across cognitive domains (all $p>0.14$). Analyses were repeated using the ranks of the 10 individual Cognition tests as outcomes, but no significant associations were found for age, sex or administration interval in the speed or accuracy domain after adjusting for multiple testing (FDR⁸).

The Cognition outcome that most strongly correlated with 6df performance was average RT on the Digit Symbol Substitution Test (DSST; Table I; model results for all 60 Cognition outcomes can be found in supplementary Table S2). In semi-partial models, DSST average RT was able to explain 34.1% of the variance in 6df performance (24.4% cross-validated). Response speed on the following tests also explained at least 10% of the variance in 6df performance (semi-partial models): Line Orientation Test (LOT), low difficulty items on the Abstract Matching Test (AM) and low difficulty items on the Matrix Reasoning Test (MRT). Signs of unadjusted correlations indicate lower 6df performance with longer response times. A high number of excess clicks on the LOT (i.e., adjustments before accepting the final line position) and Fractal 2-Back (F2B) false alarm rate were associated with lower 6df performance. High accuracy and few false starts on the Psychomotor Vigilance Test (PVT) were associated with better 6df performance. Subjects who showed a tendency to categorize faces on the Emotion Recognition Task (ERT) as angry performed better on 6df.

Using cross-validated best-subset regression, a model containing three Cognition outcomes (DSST Average RT, AM accuracy, and conservative response bias on the F2B; Table II) explained the highest amount of variance in 6df performance. The results of best subset models with 2 to 12 variables are provided in Supplement Table S2.

Discussion

This study tested how well performance on the 10 tests of the Cognition test battery predicted 6df docking performance—an operationally relevant task that simulates a spacecraft docking maneuver. Neurocognitive performance parameters were examined both individually and in models combining up to 12 outcomes. The variable most strongly correlated with 6df performance was response time on the DSST. Even after adjusting for the effects of age, sex, and administration interval on Cognition performance and after 10-fold cross-validation, DSST response time was still able to predict 24.4% of the variance in 6df performance. The DSST is a measure of processing speed requiring complex scanning,

visual tracking and working memory, and it is not surprising that these cognitive domains are important for success on a complex docking task.

The individual Cognition outcome with the second highest correlation with 6df performance was response time on the LOT. Excess clicks on the LOT were also negatively correlated with 6df performance. The LOT is a measure of spatial orientation, and it is again not surprising that performance on this test was predictive of 6df performance. While the LOT was not part of the best 3-variable prediction model, LOT accuracy or speed were part of each of the best 4 to 12 variable prediction models, stressing the importance of spatial skills for the complex operational 6df task.

Several other Cognition test outcomes were individually correlated with 6df performance. In general, higher response speed (DSST, LOT, AM low difficulty items, MRT low difficulty items) and vigilant attention was associated with higher 6df performance. Furthermore, impulsive behavior on the PVT and on the F2B were associated with lower 6df performance. Unpredictably, the tendency to score faces as “angry” on the ERT was positively correlated with 6df performance. However, the tendency to rate angry faces was also non-significantly correlated with male sex (unadjusted $r=0.224$; $p=0.2341$) and low age (unadjusted $r=-0.114$; $p=0.5478$), both variables associated with better 6df performance. Indeed, the variance in 6df performance explained by the tendency to score “angry” faces on the ERT dropped to 1.9% in semi-partial cross-validated models.

MP, VOLT and BART had no outcome that individually predicted >10% of the variance in 6df performance in semi-partial models, suggesting that sensorimotor speed, memory and risk decision making may be less important for 6df performance than the other domains covered by the Cognition battery.

Finally, we found that in adjusted and cross-validated models combining up to 12 Cognition outcomes, the model that predicted the highest variance in 6df performance consisted of 3 variables: DSST Average RT, accuracy on the AM, and bias on the F2B. Collectively, these 3 outcomes were able to predict 30% of the variance in 6df performance. Higher bias values on the F2B correspond to a more conservative response behavior (no to all), which is consistent with the finding that a lower false alarm rate was associated with higher 6df performance. AM accuracy was marginally associated with better 6df performance, but it explained only 7.9% of the 6df performance variance in semi-partial models.

The study has several strengths and limitations. The study investigated 30 high-performing individuals. While this feature increases external validity for high-performing astronauts, there may be other differences between the study participants and astronauts, and the results may not translate to other populations. That each study participant performed both Cognition and the 6df tasks repeatedly ($N=15$) is a strength of this study, but that entailed a relatively small sample size ($N=30$). This small sample size with a limited range of performance variability likely also accounts for why the explained variance in 6df performance was low for many of the individual Cognition outcomes after cross-validation. Also, this was a ground-based study that investigated participants not exposed to any of the typical spaceflight stressors. Future studies should therefore investigate associations between

Cognition and operationally relevant performance in spaceflight and in ground-based spaceflight analogs (e.g., head-down tilt bed rest, Antarctic winter-overs) to investigate whether the relationship between basic cognitive tests and operational performance differs from the findings of this study. It would also be important to investigate associations between Cognition performance and other operationally relevant tasks such as NASA's Robotic On-Board Trainer for Research (ROBoT-r).¹⁷ The latter administers similar tasks repeatedly (as opposed to the 6df training program used in this study), which allows for investigating how the association between Cognition and ROBoT-r performance changes with repeated administration.

Finally, the statistical technique used (best-subset, "exhaustive search" regression), while producing a parsimonious model, suffers from two primary weaknesses. First, the final, optimal model is a linear regression, meaning it does not account for curvilinear relationships, nor for variable-by-variable interactions. Second, the final model coefficients, while guaranteed to be the best linear unbiased estimators (BLUE),²⁴ could probably be improved by some type of regularization (e.g. ridge regression).³⁰ The effect would be minor in our case due to minimal collinearity among predictors in our final model, but it is possible for them to be improved. Overall, on the spectrum from parsimonious, interpretable models to optimally predictive models, the method used here is closer to the former. With larger samples, future research will allow improved prediction using ensembled machine learning methods.

In conclusion, this is a novel study correlating performance on basic cognitive tests with complex operational performance on a spacecraft docking task. The results suggest that cognitive throughput on the DSST – a measure of processing speed requiring complex scanning, visual tracking and working memory – was the best predictor of 6df ISS docking performance. Furthermore, high levels of spatial orientation efficiency and vigilant attention, low levels of impulsivity, and fast response speed were associated with high 6df performance. After semi-partial cross-validation, a model with three Cognition outcomes (DSST average response time, Abstract Matching accuracy, and conservative response bias on the Fractal 2-Back test) explained 30% of the variance in 6df performance. This study thus demonstrates links between performance on tests designed to assess specific cognitive domains and complex operational docking performance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The research was supported by NASA through grants NNX14AM81G, NNX14AH27G, and NNX14AH98G, by the National Space Biomedical Research Institute (NSBRI) through NASA NCC 9-58 and by ESA through grant AO-13-Concordia-13. The authors declare no conflicts of interest related to the work presented in this manuscript.

References

1. Barger LK, Flynn-Evans EE, Kubey A, Walsh L, Ronda JM, Wang W, et al. Prevalence of sleep deficiency and hypnotic use among astronauts before, during and after spaceflight: an observational study. *Lancet Neurol.* 2014;13(9):904–12. [PubMed: 25127232]
2. Basner M, Dinges DF. Maximizing sensitivity of the Psychomotor Vigilance Test (PVT) to sleep loss. *Sleep.* 2011;34(5):581–91. [PubMed: 21532951]
3. Basner M, Mollicone D, Dinges DF. Validity and Sensitivity of a Brief Psychomotor Vigilance Test (PVT-B) to Total and Partial Sleep Deprivation. *Acta Astronaut.* 2011;69(11–12):949–59. doi: 10.1016/j.actaastro.2011.07.015. [PubMed: 22025811]
4. Basner M, Savitt A, Moore TM, Port AM, McGuire S, Ecker AJ, et al. Development and validation of the Cognition test battery Aerospace Medicine and Human Performance. 2015;86(11):942–52. [PubMed: 26564759]
5. Basner M, Nasrini J, Hermosillo E, McGuire S, Dinges DF, Moore TM, et al. Effects of -12° head-down tilt with and without elevated Levels of CO₂ on cognitive performance: the SPACECOT study. *J Appl Physiol.* 2017;124(3):750–60. [PubMed: 29357516]
6. Basner M, Hermosillo E, Nasrini J, McGuire S, Saxena S, Moore TM, et al. Repeated Administration Effects on Psychomotor Vigilance Test Performance. *Sleep.* 2018;41(1):zxx187: 1–6. doi: 10.1093/sleep/zxx187.
7. Basner M, Hermosillo E, Nasrini J, Saxena S, Dinges DF, Moore TM, et al. Cognition test battery: Adjusting for practice and stimulus set effects for varying administration intervals in high performing individuals. *J Clin Exp Neuropsychol.* 2020;42(5):516–29. doi: 10.1080/13803395.2020.1773765. [PubMed: 32539487]
8. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological).* 1995;57(1):289–300.
9. Benton AL, Varney NR, Hamsher KD. Visuospatial judgment. A clinical test. *Arch Neurol.* 1978;35(6):364–7. [PubMed: 655909]
10. Dinges DF. An overview of sleepiness and accidents. *JSleep Res.* 1995;4(S2):4–14. [PubMed: 10607205]
11. Dinges DF, Pack F, Williams K, Gillen KA, Powell JW, Ott GE, et al. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 hours per night. *Sleep.* 1997;20(4):267–77. [PubMed: 9231952]
12. Furnival GM, Wilson RW. Regressions by leaps and bounds. *Technometrics.* 1974;16(4):499–511. doi: 10.1080/00401706.1974.10489231.
13. 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. [PubMed: 9345704]
14. Glahn DC, Cannon TD, Gur RE, Ragland JD, Gur RC. Working memory constrains abstraction in schizophrenia. *Biol Psychiatry.* 2000;47(1):34–42. [PubMed: 10650447]
15. Gur RC, Ragland JD, Moberg PJ, Turner TH, Bilker WB, Kohler C, et al. Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacology.* 2001;25(5):766–76. doi: 10.1016/S0893-133X(01)00278-0. [PubMed: 11682260]
16. Gur RC, Richard J, Hughett P, Calkins ME, Macy L, Bilker WB, 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):254–62. doi: 10.1016/j.jneumeth.2009.11.017. [PubMed: 19945485]
17. Ivkovic V, Sommers B, Cefaratti DA, Newman G, Thomas DW, Alexander DG, et al. Operationally Relevant Behavior Assessment Using the Robotic On-Board Trainer for Research (ROBoT-r). *Aerosp Med Hum Perform.* 2019;90(9):819–25. doi: 10.3357/AMHP.5324.2019. [PubMed: 31426899]

18. Johannes B, Salnitski V, Dudukin A, Shevchenko L, Bronnikov S. Performance Assessment in the PILOT Experiment On Board Space Stations Mir and ISS. *Aerosp Med Hum Perform*. 2016;87(6):534–44. doi: 10.3357/AMHP.4433.2016. [PubMed: 27208676]
19. Johannes B, Bronnikov S, Bubeev YA, Dudukin A, Hoermann H-J, Frett T, et al. A Tool to Facilitate Learning in a Complex Manual Control Task. *International Journal of Applied Psychology*. 2017;7(4):79–85.
20. Johannes B, Bronnikov S, Bubeev JA, Arz M, Piechowski S, Rittweger J. Individual learning curves in manual control of six degrees of freedom. *International Journal of Applied Psychology*. 2019;9(2):47–51.
21. Kane RL, Short P, Sipes W, Flynn CF. Development and validation of the spaceflight cognitive assessment tool for windows (WinSCAT). *Aviat Space Environ Med*. 2005;76(6 Suppl):B183–91. [PubMed: 15943211]
22. Kelly S. *Endurance – A Year in Space, a Lifetime of Discovery*. New York, NY: Alfred A. Knopf; 2017.
23. Lejuez CW, Read JP, Kahler CW, Richards JB, Ramsey SE, Stuart GL, 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. [PubMed: 12075692]
24. Puntanen S, Styan GPH. The Equality of the Ordinary Least Squares Estimator and the Best Linear Unbiased Estimator. *The American Statistician*. 1989;43(3):153–61. doi: 10.2307/2685062.
25. Ragland JD, Turetsky BI, Gur RC, Gunning-Dixon F, Turner T, Schroeder L, et al. Working memory for complex figures: an fMRI comparison of letter and fractal n-back tasks. *Neuropsychology*. 2002;16(3):370–9. [PubMed: 12146684]
26. Raven J The Raven's Progressive Matrices: Change and stability over culture and time. *Cogn Psychol*. 2000;41(1):1–48. doi: 10.1006/cogp.1999.0735. [PubMed: 10945921]
27. Roalf DR, Ruparel K, Gur RE, Bilker W, Gerraty R, Elliott MA, et al. Neuroimaging predictors of cognitive performance across a standardized neurocognitive battery. *Neuropsychology*. 2014;28(2):161–76. doi: 10.1037/neu0000011. [PubMed: 24364396]
28. Strangman GE, Sipes W, Beven G. Human cognitive performance in spaceflight and analogue environments. *Aviat Space Environ Med*. 2014;85(10):1033–48. doi: 10.3357/ASEM.3961.2014. [PubMed: 25245904]
29. Usui N, Haji T, Maruyama M, Katsuyama N, Uchida S, Hozawa A, et al. Cortical areas related to performance of WAIS Digit Symbol Test: a functional imaging study. *Neurosci Lett*. 2009;463(1):1–5. doi: 10.1016/j.neulet.2009.07.048. [PubMed: 19631255]
30. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2005;67(2):301–20. doi: 10.1111/j.1467-9868.2005.00503.x.

Table I:

Bivariate rank correlations between Cognition outcomes and 6df performance (ordered descending based on semi-partial R^2)

Variable	Unadjusted Spearman r	Unadjusted R^2	Semi-partial R^2	Semi-partial cross-validated R^2
DSST Average RT	-0.550 ^{**}	0.302	0.341	0.244
LOT Average RT	-0.382 [*]	0.146	0.208	0.084
AM Average RT Low Difficulty Items	-0.439 [*]	0.193	0.167	0.050
LOT Excess Clicks	-0.254	0.065	0.164	0.047
F2B False Alarm Rate	-0.434 [*]	0.189	0.145	0.052
PVT Accuracy	0.385 [*]	0.148	0.116	0.002
ERT Tendency Angry	0.478 ^{**}	0.229	0.116	0.019
MRT Average RT Low Difficulty Items	-0.211	0.045	0.108	0.002
PVT False Starts	-0.336	0.113	0.104	0.006

Semi-partial correlation adjusting for age, sex, and administration interval; DSST: Digit Symbol Substitution Test; RT: Response Time; LOT: Line Orientation Test; AM: Abstract Matching Task; F2B: Fractal 2-Back Test; PVT: Psychomotor Vigilance Test; ERT: Emotion Recognition Test; MRT: Matrix Reasoning Test; BART: Balloon Analog Risk Task;

*
p<0.05;

**
p<0.01

Table II:

6df performance predictions – model with highest cross-validated R^2

Cognition Variable	Standardized Coefficient	p-value
DSST Average RT	−0.5870	0.0005
AM % Correct	0.2960	0.0679
F2B Bias B	0.0915	0.5605

$R^2 = 0.44$; adjusted $R^2 = 0.37$; cross-validated $R^2 = 0.30$

DSST: Digit Symbol Substitution Test; AM: Abstract Matching Test; F2B: Fractal 2-Back Test