Operationally Relevant Behavior Assessment Using the Robotic On-Board Trainer for Research (ROBoT-r)

Vladimir Ivkovic; Brett Sommers; David A. Cefaratti; Graeme Newman; David W. Thomas; D. Greg Alexander; Gary E. Strangman

INTRODUCTION: Spaceflight can strain astronaut physical, physiological, and mental well-being, whereas maintaining astronaut operational performance remains an essential goal. Although various cognitive tests have been used for spaceflight assessment, these have been challenged on their lack of operational relevance.

METHODS: To address this gap, we developed and characterized the Robotic On-Board Trainer for Research (ROBoT-r) system, based on the Robotic On-Board Trainer (ROBoT) currently used for astronaut training on Canadarm2 track-and-capture activities. The task requires use of dual hand-controllers (6 degrees of freedom) to grapple an incoming vehicle in free-drift in a time-limited setting. After developing a platform for conducting research studies, characterization testing of ROBoT-r was completed by 14 astronaut-like volunteers (35 ± 11 yr; $N = 5$ women) over 16 sessions each.

RESULTS: We describe the design and capabilities of the ROBoT-r system for conducting operationally relevant research on human performance. Version 6.2 of the system supports H-II Transfer Vehicle track-and-capture operations within a multimillion component, physics-enabled 3D model using NASA's DOUG graphics platform. It has configurable task initialization and auto-run capabilities, saves 38 variables continuously at 20 Hz throughout each run, provides the user quantitative feedback after each run, and provides summaries after each session. Detailed performance characterization data is reported for future experimental planning purposes.

DISCUSSION: ROBoT-r's range of performance variables enables detailed and quantitative performance assessment. Its use in spaceflight will help provide insight into operational performance, as well as allowing investigators to compare these results with more traditional cognitive tests to help better understand the interaction between individual cognitive abilities and operational performance.

KEYWORDS: Canadarm2, robotic operation, track and capture, operational performance.

Ivkovic V, Sommers B, Cefaratti DA, Newman G, Thomas DW, Alexander DG, Strangman GE. ***Operationally relevant behavior assessment using the Robotic On-Board Trainer for Research (ROBoT-r)*****. Aerosp Med Hum Perform. 2019; 90(9):819–825.**

The extreme nature of spaceflight—and particularly long-duration (LD) exploration class missions—imposes high performance standards with small margins for error. While astronauts are carefully selected and highly trained, there is always concern about a behavioral emergency, defined by NASA as any neurobehavioral or cognitive symptoms that could result in a crewmember becoming incapacitated or cause severe mission impact. While extreme behavioral emergencies are unlikely, an impaired ability to perform key behavioral tasks remains a significant risk. Although more than 30 spaceflight studies and over 50 studies in spaceflight analogs have been performed on neurobehavioral and cognitive tasks, the evidence for alterations in LD missions remains inconclusive.[24] In large part this is because studies have been small, used different measures, and had inadequate controls, thus preventing firm conclusions one way or the other.

In addition, the majority of behavioral performance assessments used to characterize the effects of spaceflight (and

From the Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; the Center for Space Medicine, Baylor College of Medicine, Houston, TX, USA; NASA DST Technology Lab, Harmony Lane Studios, Inc., Houston, TX, USA; and NASA Flight Operations Directorate, NASA Johnson Space Center, Houston, TX, USA.

This manuscript was received for review in January 2019. It was accepted for publication in June 2019.

Address correspondence to: Dr. Gary Strangman, Ph.D., Associate Professor of Psychology, Harvard Medical School, Massachusetts General Hospital, Department of Psychiatry, 149 13th St., Charlestown, MA 02129, USA; strang@mgh.harvard.edu.

DOI: https://doi.org/10.3357/AMHP.5324.2019

analogs) are highly reductionist and do not directly translate as metrics of operational task performance.[1,11] A handful of operationally relevant tasks have been used, including timed crew efficiency,[12] Contaminants Monitoring Task (CMT),[6] Cabin Air Management System (CAMS),[22] long-arm centrifuge (LAC) manual task simulator,[3] modified Altair Lunar Lander (ALL) simulator,[2] Multi-Attribute Task Battery II (MATB-II),[21] the Robotic Workstation,[13] and SpaceDock.[25] However, with the exception of timed crew efficiency—a coarse measure of cognitive performance status—all these tests have limitations similar to cognitive tasks: they are separate tasks, disassociated from the astronaut training flow, and consequently of lesser operational relevance due to impracticality of their use in LD missions. The only operational performance assessment platform that has been both tested in spaceflight and integrated in training flows is the PILOT computerized Soyuz docking simulator.[8] A related tool is the 6df system, which is based on PILOT and also simulates manual docking of the Soyuz to a space station.[7] However, because piloting the Soyuz is strictly limited to select cosmonauts, this system has less operational and research utility for other crewmember and/or analog populations.

NASA's operationally relevant training systems potentially lend themselves to investigations relevant to a broader range of crewmember and analog populations. Systems include the Dynamic Skills Trainer (DST; advanced non-VR training platform) and the Robotic On-Board Trainer (ROBoT; derived from DST)[9]—platforms designed for training astronauts to perform various simulated capture and grappling maneuvers using the electro-motor controlled Canadarm2. The ROBoT system can simulate the critical spaceflight maneuver of capturing a free-flying spacecraft. This requires multiple cognitive processes for successful completion, including, to varying degrees, spatial visualization and orientation, situation analysis, planning, working and short-term memory, executive functioning, decision-making, object orientation, mental rotation, visual processing, fine motor control, and visual-motor integration. Such a task could provide a "yardstick" for translating the magnitude of cognitive decrements (measured by modified psychomotor vigilance tests (PVT) and the Cognition suite[1,15]) to changes in operationally relevant task performance (measured by ROBoT for Research). Importantly, proficiency with DST/ROBoT is mandatory for all astronauts who fly, making training on ROBoT an integral part of the astronaut training flow. DST and ROBoT have had limited research utility, however, because the systems have not supported research designs, data recording, or quantitative performance evaluation. This has prevented its use for behavioral performance and health investigations.

Here we describe the development of a new software tool—ROBoT for Research (ROBoT-r)—which enables deployment and detailed quantification and analysis of operationally relevant performance tests. ROBoT-r capabilities include: 1) experimental control over runs and sessions; 2) recording of all data; 3) real-time performance monitoring; 4) postrun and postsession performance evaluations; and 5) tools for quantitative data analysis. We describe the system's design and capabilities and provide effect-size characterization data in an astronaut-like cohort for future study planning. Future studies will be performed to characterize the relationship between ROBoT-r operational performance metrics and the multiple underlying cognitive processes noted above.

METHODS

Overview

The existing ROBoT system can simulate the particularly critical and difficult spaceflight maneuver of capturing a free-flying spacecraft using highly realistic 3D simulations of the arm and associated physics, and built atop NASA's Dynamic Onboard Ubiquitous Graphics (DOUG) platform.[10] The physical system (**Fig. 1A**) involves a left-hand translational controller (x/y/z) and a right-hand rotational controller (pitch/yaw/roll), plus two laptop computers. The upper laptop displays a realistic 3D rendering of the entire scene, including grappling target and arm, as viewed from multiple cameras (**Fig. 1B**). The lower laptop runs the simulation proper and displays graphs and status indicators for monitoring the simulation state and user's performance in real-time (**Fig. 1C**). Hand controllers are positioned on both sides of the simulation laptop computer at approximately shoulder width to allow unimpeded bimanual manipulation. On the ISS (0-g environment), users adopt a rough standing posture, while in the laboratory setting (1-G environment) the user maintains an upright seated posture to minimize fatigue.

Capabilities

ROBoT-r is contained within the same code base as the operational ROBoT system, but with the research recording walled off from operational use. The adaptation: 1) provides researcher control of session, grouping, and run parameters; 2) records 38 variables (**Table I**)—including user inputs, vehicle and robotic arm trajectories, and various simulation parameters—all at 20 Hz; 3) enables configurable user feedback after each run and session; and 4) generates typical "per run" behavioral metrics. For the experimenter, one can control the initial conditions (and hence run difficulty) by adjusting the starting location and attitude of the H-II Transfer Vehicle (HTV), as well as the starting linear and angular velocities. Any number of runs can be collected into groups and any number of groups can be tied together into a session. Runs within any group can be randomly selected from a larger set of options.

Metrics

Performance metrics derived from raw simulation data are based on the qualitative metrics used by astronaut trainers. These include position and orientation accuracy throughout the approach and at capture, number of capture attempts, time to completion, smoothness of approach trajectory, and collisions. To develop an overall success/failure score, variables with different units were mapped onto a 0-10 scale, as described in **Table II**.

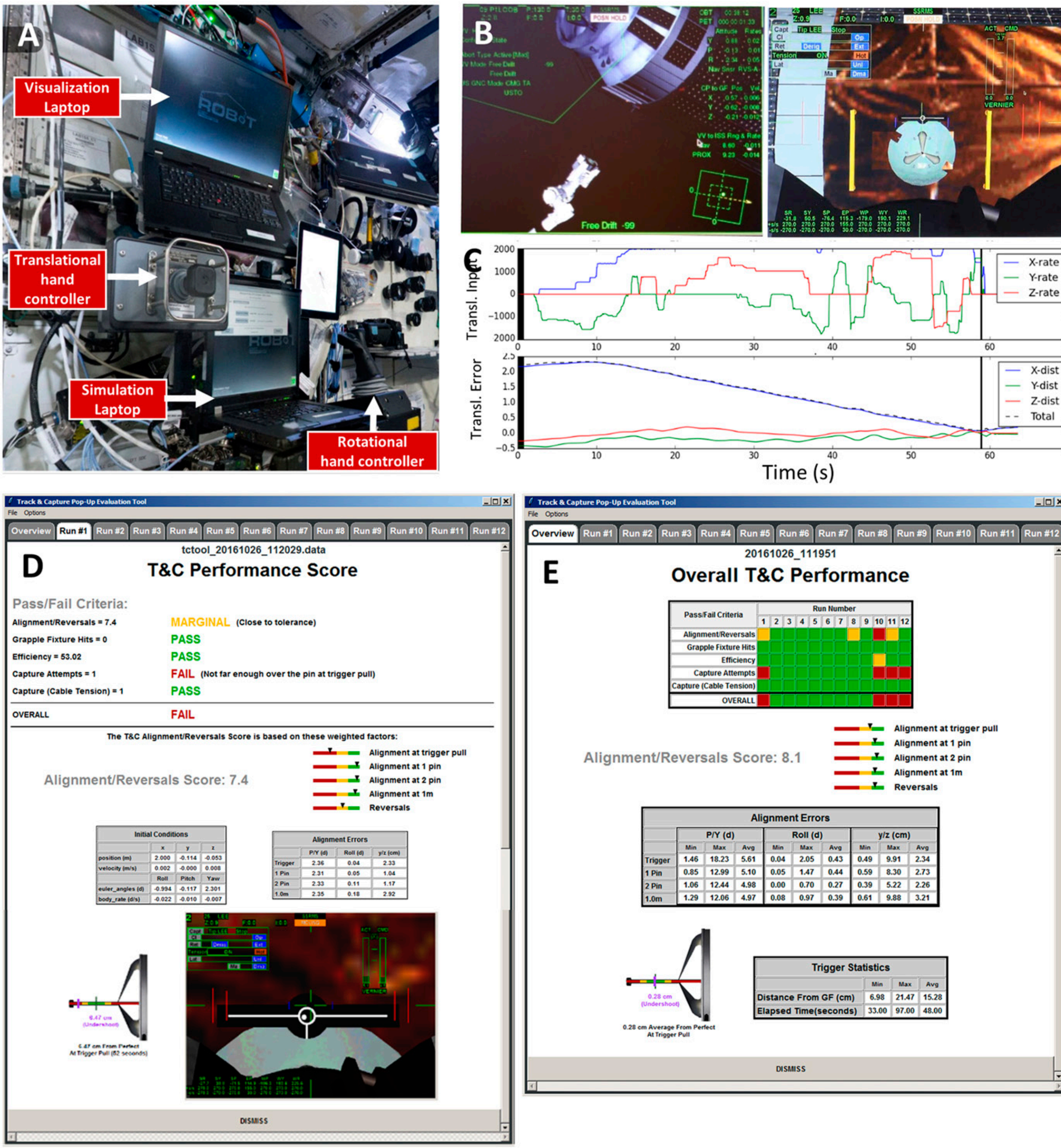

Fig. 1. A) ROBoT-r installation onboard ISS. B) Example camera views of the HTV and arm (left) and through the end-effector camera (right). C) Example translational input data (top) and resulting trajectory toward the HTV pin (bottom). D) Feedback provided after individual runs, including a pass/fail section (top), quantitative alignment scores (middle), and visual depictions of the position of the arm at the moment of capture. E) End-of-session feedback, including summary of all runs (top), average alignment errors (middle) and average timing and depth errors (bottom).

These subscores are provided as feedback to the user after each run, along with an overall alignment/reversals score which is computed by a weighted combination of translational, rotational, and reversal variables, as follows:

$$\text{alignment\_reversals\_score} = (5*\text{align@trigger}+5*\text{align@1pin}+3*\text{align@2pin}+2*\text{reversals})/15.$$

This formula more heavily weights the score toward alignment when the end effector (EE) is near or over the grapple fixture (GF) pin, while also penalizing excessive numbers of direction changes. Internally the individual scores are percentages, with 0% = worst and 100% = best, then divided by 10, making 10 a perfect score as is consistent with other astronaut evaluation scales.

Table I. Performance and Simulation Variables Recorded by ROBoT-r.

METRIC	TYPE	DETAILS
Locations	Continuous (20 Hz)	HTV grapple fixture (GF) and Canadarm2 end-effector (EE)
Alignment	Continuous (20 Hz)	x/y/z and yaw/pitch/roll of HTV GF relative to Canadarm2 EE
Movement	Continuous (20 Hz)	x/y/z and yaw/pitch/roll movement of both HTV GF and Canadarm2 EE
Hand controller input amplitudes	Continuous (20 Hz)	x/y/z and yaw/pitch/roll axis inputs
Flags	Continuous (20 Hz)	trigger pull, unfreeze dynamics, GF tip in EE, vehicle mode, cupola state, lab state, arm control mode, safing status (lee, joint, whole-arm), snare open/closed, pin contact, GF contact location (plate 0, plate1, tip, side), switches (backup, active, capture, release), cable tension (successful capture)
Time	Continuous (20 Hz)	Current (computer) time stamp, duration since run start
Performance	Per run (both quantitative and pass/fail)	Overall, alignment/reversals, grapple fixture hits, efficiency (elapsed time), capture attempts
Initialization conditions	Per run	Starting location/orientation of HTV, HTV translational/rotational velocities, runID, pin length

HTV: H-II transfer vehicle.

Defaults

Within the above framework, ROBoT-r has also been given a default configuration. In this configuration, there are 400 initial conditions, grouped into 4 difficulty levels, 100 per level. A standard testing session involves 12 runs (i.e., grappling attempts), with 3 runs randomly selected from each of the 4 difficulty levels. Difficulty is defined as a percentage of the maximum values allowed for a grappling attempt (e.g., the maximum translational and angular displacement and motion the HTV can have for Mission Control to allow a capture attempt). The four default difficulty levels are 25%, 50%, 75%, and 100% of these maximum levels, where maximum x/y/z velocity is 0.033 $m \cdot s^{-1}$ (root sum of squared deviations; RSS) and maximum angular velocity is $0.1° \cdot s^{-1}$ (RSS), with initial positions of $x = 2.0$ m, y/z offsets < 0.5 m (max for operational training $=$ 0.16 m), and pitch/roll/yaw misalignments $< 10°$ (max for operational training $= 4°$). The 12 runs are presented in 4 steps of progressively increasing difficulty. Runs are an automatic failure after 99 s, with any attempt that takes > 80 s considered marginal. In all, a standard 12-run session takes approximately 25 min to complete.

Table II. Scoring Protocol for Each Performance Variable.

VARIABLE	SCORING
X (depth)	0.145–0.155 m from base plate scores a 10
	≥ 0.1 m error from 0.15 m scores a 0
	Linear in between
Y, Z (in-plane)	≤ 0.005 m (root sum of squared deviations; RSS) scores a 10
	≥ 0.16 m RSS scores a 0
	Linear in between
Roll	≤ 0.5° scores a 10
	≥ 16° scores a 0
	Linear in between
Pitch, Yaw	≤ 0.5° RSS scores a 10
	≥ 23° RSS scores a 0
	Linear in between
Reversals	0 reversals across all axes score a 10
	≥ 20 reversals score a 0
	Linear in between
Grapple fixture hits	0 hits score a 10
	1 hit scores a 5
	2 or more hits score a 0
Efficiency	0–60 s scores a 10
	≥ 100 s scores a 0
	Linear in between
Capture attempts	0 attempts score a 0 (failed to try)
	1 attempt scores a 10
	2 attempts score a 5
	3 attempts score a 0
Cable tension	1 (tension) scores a 10 (pin captured)
	0 (no tension) scores a 0 (pin not captured)

Training

Standard training procedures for new users involve a 30-min orientation to the system, a 30-min session involving one-on-one demonstration plus minimal hands-on use, and then two 30-min hands-on training sessions—the first with ample instructor feedback highlighting the feedback criteria, and the second where instructor feedback is only provided when requested or for criteria on which the individual receives a failing evaluation. Users are taught to minimize translational and rotational errors, minimize reversals of direction along any of the six axes, minimize capture attempts, and avoid bumping the vehicle with the arm.

Implementation

To start each ROBoT-r session, subjects log in and start an auto-run session. The first run is then initialized (requiring ~100 s on a HP Zbook 15 G2), five beeps sound, the user initiates the run by "unfreezing" the simulation dynamics, and a grappling attempt is made. The run ends either when cable tension is achieved, or 99 s has elapsed, whichever comes first. Immediately after the run a feedback screen is displayed for that run (**Fig. 1D**) and the next run is initialized. After the final run, a summary feedback screen is displayed that summarizes performance from all 12 runs from the current session (**Fig. 1E**). Comparison of performance to baseline or previous test sessions is currently only possible offline, not from within ROBoT-r simulation.

Characterization

To characterize the task, we recruited a total of $N = 17$ healthy demographically astronaut-like subjects holding a Master's degree or higher in STEM fields or relevant technical or military equivalent experience, between the ages of 30 and 55 yr. The subjects did not have prior experience with spaceflight control systems. Each performed the ROBoT-r task in our laboratory at Massachusetts General Hospital (MGH), on a 7-wk schedule that included (after training), 2 baseline data collection sessions, 12 experimental sessions at 1–4 d intervals spread over 30 d, plus 2 postexperimental sessions using the default run configuration. ROBoT-r recorded an overall score plus 38 distinct performance metrics (Table I). Behavioral results were analyzed via mixed effects linear regression analysis[18] clustering by subject.

RESULTS

Data from $N = 14$ (ages 35 ± 11 yr; 5 women) were analyzed (2 subjects were excluded from the study due to noncompliance with instructions, and 1 subject due to an incomplete dataset). **Fig. 2** shows data from the 16 consecutive sessions performed following subjects' initial basic training as described above. Overall performance (weighted scores) improved significantly ($z = 8.8$, $P < 0.001$, slope = +0.10/session) while time to completion significantly decreased ($z = -7.4$, $P < 0.001$, slope = −0.70 s/session; see Fig. 2A). Together, these findings are consistent with continued learning of the highly complex capture task over the 16 sessions (approximately 8 total hours) of performance. We tested both log and exponential fits to the data, but the linear model produced a better fit (BIC criteria). When looking across runs within sessions (Fig. 2B), one sees a clear stepped structure related to run difficulty, both in overall performance (run: $z = 2.0$, $P = 0.048$, slope = +0.071/run; difficulty: $z = -9.1$, $P < 0.001$, slope = −0.04/run) as well as time to completion (run: $z = -4.9$, $P < 0.001$, slope = −1.4 s/session; difficulty: $z = 8.9$, $P < 0.001$, spanning 15 s from easiest to hardest). Means and standard deviations in key performance metrics are included in **Table III** for future experimental planning purposes. Note that even after 16 sessions, the mean overall success proportion is 0.52—that is, just over 50% of trials were successful on all graded metrics—highlighting the challenge of the task. In structured postexperiment subject debrief interviews, 96% of subjects reported the task as challenging and consistently motivating.

DISCUSSION

ROBoT-r provides a unique, highly realistic, and operationally relevant performance test that can be used for spaceflight and ground-analog research testing. The default configuration provides a range of difficulty levels that are clearly challenging yet motivating—an important feature for behavioral testing. The range of output variables are suitable for identifying learning effects, enabling quantitative investigation of departures from nominal performance levels, and identifying various operational deficit domains (e.g., collisions with the HTV, performance slowing, nonsmooth or "weaving" approach trajectories, and potentially changes in strategy). Given the detailed raw

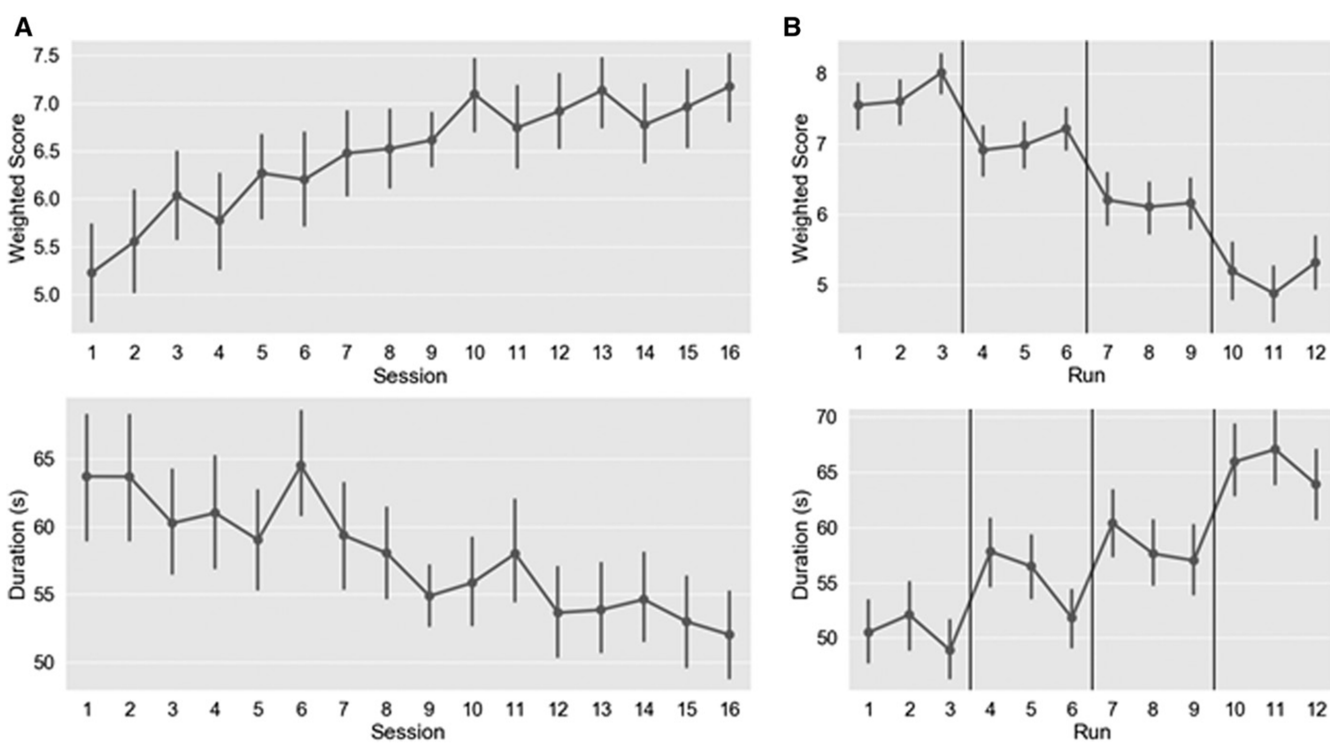

Fig. 2. ROBoT-r performance. A) Weighted performance score (top) and run duration (bottom) over 16 sessions ($N = 14$). B) The same two outcomes plotted as a function of run within a session. Stepped structure and vertical lines identify different run difficulty levels ($N = 14$). Error bars represent bootstrapped 95% confidence intervals.

Table III. Performance at Session 1 vs. Session 16.

OUTCOME	SESSION 1 MEAN (SD)	SESSION 16 MEAN (SD)	MEAN DIFF.	PERCENT CHANGE
Weighted score	5.2267 (2.9866)	7.1744 (2.2896)	1.9477	37.3
Run duration (s)	63.6991 (26.6362)	52.0352 (19.6103)	−11.6639	−18.3
Pin capture (%)	82.5 (38.156)	96.7949 (17.6704)	14.2949	17.3
Distance error (m)	0.2083 (0.2599)	0.1705 (0.2162)	−0.0378	−18.1
Angle Error (degrees)	6.9295 (6.095)	4.327 (3.6584)	−2.6026	−37.6
GF contact	1.0917 (2.6186)	0.2949 (0.9852)	−0.7968	−73.0
Trigger pulls	1.1917 (0.639)	1.0192 (0.2402)	−0.1724	−14.5
Overall pass/fail (1/0)	0.175 (0.3816)	0.5192 (0.5012)	0.3442	196.7

GF: grapple fixture.

data saved from each simulation, the default performance metrics can be supplemented via additional analyses to track, for example, differences in translational vs. rotational hand controller use (rotational corrections are typically harder to make[16,25]), the number of reversals of direction or translational/angular errors separated by axis (rather than lumping all axes together), runs that subjects elected not to attempt, and so on.

A key question to consider is what constitutes a meaningful change in ROBoT-r performance. For astronauts, mission-critical maneuvers such as vehicle capture require essentially 100% success rates. Thus, any dip below 100% overall success would be considered a meaningful or concerning decrement in that population. Our subjects did not reliably achieve 100% success, however, and our experimental findings demonstrated continued learning through ~8 h of task performance, as well as substantial between-subject variability. This makes a strict threshold inappropriate for identifying "concerning" behavior in this group. One approach is to consult Table III, where changes exceeding 2 standard deviations would represent a typical threshold for significant performance decrement. Given the substantial interindividual variability, however—common in behavioral tests—it may be more powerful in some cases to evaluate individual subject performance based on their own changes over time, rather than comparing to group means,[13,20] as done for cognitive performance in the recent NASA Twins study.[4]

Prior to ROBoT-r, various other systems have been used for operationally relevant performance research, several of which include highly realistic interfaces (LAC,[3] ALL,[2] CMT,[6] CAMS[22,23]). However, most exhibit one or more of the following limitations. For example, the LAC and ALL simulators are less appropriate for use in spaceflight as the former is physically large and the latter simulates the discontinued Altair Lunar Lander. In addition, the LAC and MATB-II[21] simulators are limited to single-hand, 3-axis motion control, while the MATB-II, ALL and SpaceDock[25] systems are lower-fidelity or have few or exclusively intermittent behavioral metrics. Most important, none of these are integrated into astronaut training or performance flows. The PILOT and 6df simulators are exceptions,[7,8] providing a high-fidelity training platform and monitoring 12 distinct performance parameters. However, they are most operationally relevant for the narrow group of crewmembers that pilot the Soyuz.

As with many other complex tasks,[13,14,20] learning effects are prominent with ROBoT-r, and such effects have the potential to obscure behaviorally relevant performance impairments. This cannot be avoided entirely, but there are various approaches to consider. One is to pretrain subjects to some criterion on the task to reduce or eliminate further learning effects. This is effectively the situation with astronauts, who typically perform a minimum of 100 h of training with the ROBoT system. We thus expect ROBoT-r to be more sensitive to changes in operational performance in this group. For laboratory studies, pretraining to this extent may or may not be feasible. For studies in analogs such as NASA's Human Exploration Research Analog (HERA) or Antarctica, such pretraining is often precluded for pragmatic or availability reasons. In such cases, one can fit a suitable (e.g., logarithmic) curve to individual-subject data as an estimate of their expected learning curve[19]—and then investigate off-nominal performance after adjusting for this curve. This approach of course embeds assumptions that may not be justified—such as assuming a particular shape of monotonic learning on a task that can lead to plateaus or even periods of performance worsening during the learning process,[5] thus encouraging interpretive caution. A third approach—and partial solution for analog studies—is to include matched control groups. This is key for any group-level interpretation, although is less effective for assessing performance by individual subjects. Combining ROBoT-r performance with physiological monitoring may also help differentiate typical learning effects from other acute or chronic performance anomalies.

With ROBoT-r we sought to provide an operationally relevant performance assessment tool that is complementary to traditional cognitive testing batteries while minimizing obtrusiveness. As with all such tests, however, consistency in deployment is important. If multiple studies were to use different "versions" of ROBoT-r—e.g., using different numbers of runs per session, or different difficulty levels, feedback screens, scoring algorithms, training procedures, or adding secondary tasks—comparison of results across studies would be less reliable. This is a particular concern for spaceflight and analog research, where the data collection opportunities can be limited and *N*s are usually small.[24] Appropriate controls for astronaut subjects are also an issue, particularly given the difficulty in replicating the extensive training astronauts receive on the ROBoT system. One possibility is to use nonflying astronauts as controls, but this is not always feasible. Another limitation relates to a key feature of ROBoT-r: its operational integration. Research examining performance over time typically controls exposure to the testing task. Such control is not possible in flight with ROBoT-r, as the same system is used to practice operational track-and-capture maneuvers on an as-needed basis. This leads not only to extra exposure to the system, but can also involve exposure to different versions of the task (e.g., non-HTV capture activities). While unavoidable, such exposures should have lesser effects in highly trained individuals and can potentially be adjusted for during modeling. Finally, ROBoT-r simulates only a single type of operational task. While ROBoT-r requires many cognitive, performance, and executive faculties, qualitatively different types of operational tasks (e.g., wayfinding in low-cue environments[17]) may be differently affected by the rigors of spaceflight. A task that is broadly relevant to all operational activities has yet to be developed.

In sum, the development and characterization of ROBoT-r represents a first step toward an objective, operationally integrated research tool for assessment of behavioral health and performance for human spaceflight and ground-based analogs. Future ROBoT-r studies are expected to include investigations of the effects of mission duration, the effects of sleep status, and comparison with cognitive test batteries to help map individual cognitive changes (e.g., in spatial orientation, working memory and attention, sensorimotor ability and psychomotor speed, risk decision making) to changes in ROBoT-r performance. This approach will enable investigation of the effects of microgravity, oxygen concentrations, and other space environmental conditions and life support systems on ROBoT-r operational performance.

ACKNOWLEDGMENTS

This study was supported by NASA grant NNX15AK76A. Special thanks go to Dr. David Dinges, lead of the Behavioral Core Measures project under which this system was developed. Thanks also go to NASA managers L. Leveton, T. Williams, and J. Schneiderman for their supporting roles, as well as ROBoT project managers P. McCartney and M. Daley, and early ROBoT contractors S. Killingsworth and E. Paddock for their support and technical development efforts.

Authors and affiliations: Vladimir Ivkovic, M.Sc., Ph.D., and Gary E. Strangman, Ph.D., Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; Brett Sommers, B.Sc., David A. Cefaratti, B.Sc., and D. Greg Alexander, B.Sc., NASA DST Technology Lab, Harmony Lane Studios, Inc., Houston, TX, USA; Graeme Newman, B.Sc., Robotics, and David W. Thomas, M.S., NASA Flight Operations Directorate, NASA Johnson Space Center, Houston, TX, USA; and Gary E. Strangman, Ph.D., Center for Space Medicine, Baylor College of Medicine, Houston, TX, USA.

REFERENCES

1. Basner M, Savitt A, Moore TM, Port AM, McGuire S, et al. Development and validation of the Cognition test battery for spaceflight. Aerosp Med Hum Perform. 2015; 86(11):942–952.
2. Bilimoria KD. Effects of control power and guidance cues on lunar lander handling qualities. J Spacecr Rockets. 2009; 46(6):1261–1271.
3. Clark TK, Newman MC, Merfeld DM, Oman CM, Young LR. Human manual control performance in hyper-gravity. Exp Brain Res. 2015; 233(5):1409–1420.
4. Garrett-Bakelman FE, Darshi M, Green SJ, Gur RC, Lin L, et al. The NASA Twins Study: a multidimensional analysis of a year-long human spaceflight. Science. 2019; 364(6436). pii: eaau8650.
5. Hauptmann B, Reinhart E, Brandt SA, Karni A. The predictive value of the leveling off of within session performance for procedural memory consolidation. Brain Res Cogn Brain Res. 2005; 24(2):181–189.
6. Hockey GR, Wiethoff M. European isolation and confinement study. Cognitive fatigue in complex decision-making. Adv Space Biol Med. 1993; 3:139–150.
7. Johannes B, Bronnikov S, Bubeev Y, Dudukin A, Hoermann HJ, et al. A tool to facilitate learning in a complex manual control task. Int J Appl Psychol. 2017; 7(4):79–85.
8. 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):534–544.
9. Johnson G, Alexander G. Robotics On-Board Trainer (ROBoT). Houston (TX): NASA Johnson Space Center; 2013. Report No.: MSC-25005-1.
10. JSC TVRLVN. Dynamic Onboard Ubiquitous Graphic (DOUG), 1.72 ed. Houston (TX): The Virtual Reality Laboratory (VRlab); 2018.
11. 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–B191.
12. Kubis JF, McLaughlin EJ, Jackson JM, Rusnak R, McBride GH, Saxon SV. Task and work performance on Skylab missions 2, 3 and 4: time and motion study–experiment M151. In: Johnston RS, Dietlein LF, editors. Biomedical results from Skylab. Washington (DC): NASA Lyndon B. Johnson Space Center; 1977:136–154.
13. Liu AM, Oman CM, Galvan R, Natapoff A. Predicting space telerobotic operator training performance from human spatial ability assessment. Acta Astronaut. 2013; 92(1):38–47.
14. Liu JC, Verhulst S, Massar SA, Chee MW. Sleep deprived and sweating it out: the effects of total sleep deprivation on skin conductance reactivity to psychosocial stress. Sleep. 2015; 38(1):155–159.
15. Moore TM, Basner M, Nasrini J, Hermosillo E, Kabadi S, 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.
16. Mueller E, Bilimoria K, Frost C. Effects of control power and inceptor sensitivity on lunar lander handling qualities. J Spacecr Rockets. 2011; 48(3):454–466.
17. Newman DJ, Lathan CE. Memory processes and motor control in extreme environments. IEEE Trans Syst Man Cybern C Appl Rev. 1999; 29(3):387–394.
18. Pinheiro JC, Bates DM. In: Chambers J, Eddy W, Härdle W, Sheather S, Tierney L, editors. Mixed-effects models in S and S-Plus. New York (NY): Springer; 2000.
19. Pusic MV, Boutis K, Pecaric MR, Savenkov O, Beckstead JW, Jaber MY. A primer on the statistical modelling of learning curves in health professions education. Adv Health Sci Educ Theory Pract. 2017; 22(3): 741–759.
20. Richards JT, Oman CM, Shebilske WL, Beall AC, Liu A, Natapoff A. Training, transfer, and retention of three-dimensional spatial memory in virtual environments. J Vestib Res. 2002–2003; 12(5–6):223–238.
21. Santiago-Espada Y, Myer RR, Latorella KA, Comstock JR. The Multi-Attribute Task Battery II (MATB-II) software for human performance and workload research: a user's guide. Hampton (VA): National Aeronautics and Space Administration, Center LR; 2011. Report No.: NASA/TM–2011-217164.
22. Sauer J, Hockey GR, Wastell DG. Maintenance of complex performance during a 135-day spaceflight simulation. Aviat Space Environ Med. 1999; 70(3, Pt. 1):236–244.
23. Sauer J, Hockey GR, Wastell DG. Performance evaluation in analogue space environments: adaptation during an 8-month Antarctic wintering-over expedition. Aviat Space Environ Med. 1999; 70(3, Pt. 1): 230–235.
24. Strangman GE, Sipes W, Beven G. Human cognitive performance in spaceflight and analogue environments. Aviat Space Environ Med. 2014; 85(10):1033–1048.
25. Strangman G, Thompson JH, Strauss MM, Marshburn TH, Sutton JP. Functional brain imaging of a complex navigation task following one night of total sleep deprivation: a preliminary study. J Sleep Res. 2005; 14(4):369–375.