3.6

Spaceflight Cognitive Assessment Tool for the Lunar-Mars Life Support Test Project Phase III Test

Christopher Flynn, M.D., F.S., Daniel Eksuzian, Steven Vander Ark, Walter Sipes, Ph.D.

SUMMARY

Main Objective: To test a computer-based objective cognitive assessment tool in an analogue environment comparable to a space station.

A cognitive test was developed and tested. The Lunar-Mars Life Support Test Project (LMLSTP) Phase III test crew and back-up crew were briefed on the test, took the test four times for baseline data, and then took the test three times during a 91-day chamber test. The test was evaluated in terms of adequacy, training, scheduling, administration, and problems.

PSYCHOLOGY (COGNITIVE) SELF-EVALUATION

Introduction

Specific Aims and Objectives

For the purposes of the LMLSTP Phase III Test, the Behavioral Health and Performance Group (BHPG) assessed a cognitive assessment tool to assist in monitoring crew health during the 91-day stay in a sealed chamber. The Spaceflight Cognitive Assessment Tool (S-CAT) was designed as part of the behavioral medicine monitoring efforts to be used on the International Space Station (ISS). This tool will be incorporated into the routine medical monitoring regimen being conducted by NASA Johnson Space Center Medical Operations. The S-CAT was used to provide the medical support personnel with a valid measure that would help them make decisions regarding the cognitive well being and capabilities of the crew. In addition to routine health monitoring, this test could be used in case of accident, injury, or exposure to off-nominal environmental conditions.

There were two specific aims for the Phase III test:

- Evaluate the sensitivity and implementation of the S-CAT during a space-analogue mission; and
- Provide objective measures of cognitive functions that could be used in caring for the crew during the test.

The original intended objective of the Behavioral Health and Performance Group's LMLSTP Phase III effort was, in addition to evaluating the S-CAT, to assess a Behavioral Medicine Crew Assessment Battery. This battery was to include stress and mood assessment tools. Due to time and funding constraints, only the S-CAT was employed for this test.

Background

As of the end of 1996, there was no objective measure of cognitive functions available to space crews, although anecdotal reports from space crews suggested the space environment might adversely affect crew cognitive performance. crewmembers from short- and long-duration missions reported mild degradations in their ability to remember tasks and to recall information. Off-nominal conditions including accidents, injuries, and exposures to toxins can certainly affect an individual's ability to function. Operations aboard Space Station Mir clearly indicated the need for some form of objective cognitive and performance assessments.

Description of the S-CAT

A small team of extramural experts was assembled to assist the NASA Medical Operations BHPG develop a tool that could be used in the evaluation of the cognitive functions of space crewmembers within the time and environmental constraints of space missions. Expertise on this team included three clinical neuropsychologists and two experts in the construction and use of automated psychological tests.

Operational constraints significantly affected the development team's efforts in identifying the appropriate cognitive assessment tool. The constraints included: available computing equipment onboard ISS; the limited time available for completing the tool; results having to be immediately fed back to the crewmembers; and given the critical need and the short timeframe for development, the tool by necessity would be created from existing, validated tests and methods.

The S-CAT construction was based on a battery of tests derived from the Automated Neuropsychological Assessment Metrics (ANAM) developed for the U.S. Department of Defense (3, 4). The ANAM consists of validated tests that have been used in clinical settings to evaluate personnel with suspected brain injury (3). The ANAM tests used in the S-CAT were selected to meet the time and diagnostic requirements of the space environment. The following is the list of the S-CAT subtests and a brief description of what each measures:

- Code Substitution–memory
- Running Memory–sustained concentration
- Math-verbal working memory
- Match-2-Sample–visual working memory
- Code Substitution (Recall)–recall

The development team developed both a short and long version of the S-CAT. The short version, which was developed, required approximately 15 minutes to complete, and allowed routine monitoring of crew cognitive functions and provided initial diagnostic information in the event of an injury or toxin exposure. The long version was intended to provide an enhanced diagnostic capability (using additional tests) of the crewmember's condition, if necessary, and could be completed in 35 minutes.

Methods

Protocol

To implement the S-CAT, crewmembers had to learn and become proficient at each of the tests that comprise this tool. Crewmembers were first given a short familiarization briefing and documentation about the purpose and use of the S-CAT and then they completed four sessions taking the short version and two sessions of the long version. To reduce scheduling problems, the crewmembers attended training and baseline data collection sessions in groups of four. The team of S-CAT developers estimated that at least four sessions would be needed to overcome learning effects and to produce meaningful baseline data for each crewmember.

During the test, the same protocol planned for the Space Station Mir and the ISS was followed. Each crewmember was scheduled to take at least the short version of S-CAT once per month, coinciding with the monthly physical examination, for a total of three scheduled sessions during the test. Although the crew would be reminded when to take the S-CAT, they were responsible for actually scheduling and doing the S-CAT. The long version was scheduled at approximately test days 45 and 75 for a total of two times during the Phase III test. Following completion of the Phase III test, crewmembers would be debriefed and, if possible, take the short version of S-CAT one more time.

Since a back-up crewmember could be a replacement into the test chamber at any time, the back-up crew was trained and performed baseline S-CAT testing. Thus, all potential crewmembers were proficient and had recent baseline data.

Repeatedly taking the S-CAT ensured two objectives: there would be good data available from pretest and nominal operations over time with which to compare test results following a mishap; and crewmembers would maintain proficiency on the tests so that learning (or relearning) effects would not confound the data and lead to a misdiagnosis of the crewmember's condition. Table 3.6-1 is the schedule of S-CAT sessions for the Phase III test.

Table 3.6-1 *Scheduled S-CAT sessions for prime and backup crewmembers*

Pretest	In-test	Post-test
1 Familiarization @ 1 hour	Short @ 15 minutes on test days 30, 60, and 88 and as needed	1 Debrief @ 1 hour
4 Training_short @ 30 min	Long @ 35 minutes on test days 45 and 75 and as needed	1 Short @ 15 minutes < 30 days post-test
2 Training_long @ 35 min (combined with last two short sessions)		

Hardware

The hardware used during this project consisted of DOS-based notebook and desktop computers. The make and model of the computers used are unimportant except to note that they used a DOS-based (versus Apple) operating system that is required to run the S-CAT. The familiarization and initial exposure to the S-CAT was conducted at a training classroom with desktop computers in Building 12 at Johnson Space Center. Training sessions and the baseline data collection were completed at the Krug Life Sciences (now Wyle Laboratories) Parsec II building using notebook computers. During the Phase III test, the prime crew used their personal desktop computers in the test chamber and the back-up crewmembers used a notebook computer stored in the test control room.

Additional Issues

In addition to providing a tool to assist in evaluating crew capabilities during the Phase III test, an objective of this project was to answer three important questions regarding the implementation of the S-CAT.

S-CAT Practice and Baseline Sessions

First, what amount of training and practice is required to alleviate task learning effects? Based on experience using the ANAM tests, the S-CAT development team predicted that 5 to 10 sessions would be necessary to ensure that the subject learned the tasks well enough to work at his/her best performance level. However, the Phase III crewmembers had considerably less time than that available as they prepared for the test. The result was that crewmembers had just the one familiarization meeting and the four sessions for training and baseline data collection.

Control of S-CAT Data

The second question is what is the best way to control the S-CAT data? Significant results of poor performance on a test like the S-CAT, indicating a neuro-cognitive problem, is likely to be perceived by flight (and test) crewmembers as very serious and potentially career jeopardizing. This bias is believed even if the data would not be used to reach conclusions independent of other data. Hence, there is a reluctance on the part of the crewmembers to take such tests or to share the data with the flight surgeon, medical monitor, or flight managers.

Since the S-CAT cannot help the crew or medical personnel if it is not used, then the top priority was to encourage crewmembers to use the tool. The approach applied during the Phase III test was to allow the crewmember to control use of the data. In other words, if the crewmember did not want to share the data, then they did not. With this approach, the data collected was for the crewmember alone to assess his or her own capabilities without the fear of detrimental judgment from management or the flight surgeon/medical monitor. The counter argument to this approach was that crewmembers might not recognize impairment given the limited training on interpreting the performance scores. And, if they did, they might not report it to the flight surgeon/medical monitor for fear of some reprisal either immediately or upon completion of the mission. Ultimately, it was decided that the best option was to trust the crewmember to report any anomalies, although the crew surgeon could request that the crewmember take and report S-CAT results if test events warranted.

S-CAT Schedule

Similar to the question about data control was how often should the S-CAT be scheduled? The S-CAT would be used following an event that might have resulted in cognitive impairment of crewmembers. To determine whether or not a crewmember's cognitive abilities are different from normal, the post-event S-CAT test scores would be compared to S-CAT scores produced both before the mission and during the mission up to that point. However, the crewmember must be proficient at taking the tests in order for the scores to be truly meaningful. If the crewmember fails to maintain proficiency on the S-CAT, then performance degradations due to loss of skill over time might be interpreted incorrectly as a neuro-cognitive problem.

When should the crewmember take the S-CAT to maintain proficiency? The S-CAT development team was not certain of the maximum time interval between S-CAT administrations that could occur before problems were encountered. For space crews, one approach to the problem is to couple the S-CAT with other medical requirements. Since the S-CAT is to be a medical requirement, it is logical to incorporate it with the physical examinations that occur monthly. Based on the experience of the S-CAT development team with the ANAM tests, taking the test every thirty days or so was thought to be sufficient to maintain S-CAT proficiency.

Another approach to the S-CAT schedule question was to inform the crew of the importance of taking the S-CAT periodically and letting them determine the specific schedule. On long-duration missions, especially when communications are disrupted or have lengthy delays, crews will have to act fairly independently from their Earth-based support group. In light of this more autonomous mission, perhaps the crew should set and follow a schedule of self-examination, including the S-CAT, which fits their respective work schedules. This is the approach taken for the Phase III test. While the crewmembers were briefed on the purpose of the S-CAT and the need to maintain proficiency, they were given a suggested schedule instead of hard dates on which to take the S-CAT. They were requested to complete the S-CAT around test days 30, 60, and 90 (officially, around day 88 so to avoid the exit day flurry of activities) and they were reminded when those dates approached. It is important to note that history suggests that if a task is not hard-scheduled into the extremely busy schedule of an astronaut, then it is not likely to be accomplished.

Results

Anomalies

There were some imperfections in the S-CAT software that caused the long version to be troublesome throughout the test. The training sessions were completed and several of the crewmembers attempted the long version during the test, but no meaningful baseline or in-test data were collected. The imperfections in the S-CAT software that caused the long version problems and other software-related issues have resulted in significant improvements being made to the S-CAT, including reduced conflict with a variety of computer platforms, easier installation, and fixing a previously unidentified data scoring error.

The group training that occurred for the Phase III test was not optimal. Each crewmember (or test subject) should be trained in a private area that is relatively free from distraction. Though the group setting did not seem to affect the last two pretest data collection sessions, the first two were probably not representative of what would occur in a more private setting. It is clear from direct observations and the numerous comments from the crewmembers that a requirement of using the S-CAT needs to be a fairly quiet location free of interruptions. The group sessions proved to be too distracting. For example, as one individual asked a question or commented, the others naturally stopped to listen or elaborate instead of continuing their session uninterrupted. Additionally, it was all but impossible for the back-up crewmembers to take the S-CAT in the control room due to the constant distractions of on-going communications, problem solving, and visitors.

Objective 1: Evaluate S-CAT Implementation

Training Time

For each of the prime and back-up crewmembers, performance asymptote on the S-CAT short version was reached after the familiarization session and two data collection/training sessions. If or when performance leveled off for the long version of the S-CAT is inconclusive due to the software problems mentioned above. It seems as though the number of sessions required to achieve proficiency on the S-CAT might not be as much as the S-CAT developers originally thought. Training time of future subjects will be noted to compare with the results of the Phase III test. Training for the S-CAT for future use will be one familiarization and five baseline sessions.

Efficiency

The S-CAT appeared to be a reliable measure of cognition based on preliminary examination of the data. The S-CAT development team has recommended specific criteria for go/no-go decisions. Further assessment of the effectiveness of the tool is warranted with validation studies.

Data Control

There were no untoward events (head trauma, exposure to toxins) during the mission requiring cognitive ability evaluation so evaluation of the control of data and sharing of it with the medical monitor was not tested.

Schedule

Neither the short nor the long versions of the S-CAT were taken following the recommended schedule and some crewmembers did not complete either version the recommended number of times. Those who did complete the S-CAT did not follow any schedule. When they were reminded that it was time, they attempted to complete an S-CAT test, but did not necessarily accomplish fitting it into their schedule. One or two of the back-up crewmembers attempted to take the short version once or twice, but there was no meaningful use of, or data collected from, the S-CAT by the back-up crewmembers.

Based on the nearly unanimous recommendation from the prime and back-up crewmembers and on the history of manned space operations, the S-CAT sessions must be a requirement in the astronaut's flight timeline schedule, not just a recommendation. Otherwise, the probability that the S-CAT will be taken and, therefore, be a meaningful measure when needed is low. To paraphrase the gist of what the Phase III crewmembers reported, 'There will always be something more pressing to take care of or something more desirable to do than to spend the time doing a test that will not likely be used anyway.' The importance of this type of data to crewmembers and medical personnel is critical in the case of a cognitive event.

Objective 2: Provide Objective Measures of Cognitive Functions

There were no untoward events (head injury, exposure to toxins) that warranted assessment of any crewmember's cognitive functions.

Discussion

The S-CAT will meet the on-orbit cognitive assessment needs for future space missions, but it must be a hard-scheduled medical requirement rather than a suggested tool with a recommended schedule. Further, the S-CAT should be scheduled during the normal work day to reduce time and energy conflicts in order to ensure the crewmember is "working" at full, nominal levels rather than tired and rushed at the end of the day.

The issues over control of the S-CAT data were untested in the Phase III test. Control over the data is to remain with the crewmember unless an untoward event occurs. Depending on the event, the data may be used by the crew surgeon for medically-based diagnostic and management recommendations.

The long version was not perfected for the Phase III test, so no meaningful data were collected. Even so, the lessons learned regarding the training, baseline data collection, and implementation of the long version were quite valuable. These lessons were used in preparation for the NASA-7 mission and ISS implementation.

Further development of the S-CAT continues with improvements to data presentation displays, multiple language capability, installation improvements, diagnostic criteria, selection of different tests (as deemed necessary), and compatibility with the Microsoft Incorporated Windows95® operating system.

SIGNIFICANCE

A computer-based objective cognitive assessment tool was successfully developed and tested in a 91-day chamber study to be used on Space Station Mir and the ISS. The lessons learned from this chamber study included giving a thorough briefing on the importance of this tool, hard-scheduling training/baseline and operational use of this tool, and having the crewmembers maintain control over their own data unless there is a medical event.

Future research directions include validation studies with both normal and clinical populations. Possible operational applications may include the military, physicians, underwater divers, or other high-risk occupations.

REFERENCES

1. Levinson DM, Reeves DL. Monitoring recovery from traumatic brain injury using Automated Neuropsychological Assessment Metrics (ANAM V1.0). *Arch Clin Neuropsychol (1997) 12:155-166.*
2. Reeves DL, Levinson DM, Gamache G, Bidiouk P. Neurocognitive & Physical Abilities Assessments Ten+ Years (1995-'96-'97) after the Chernobyl Nuclear Accident. National Cognitive Recovery Foundation Technical Report *NCRF-TR-98-01, 1998.* NOTE: This reference is not cited in the text.

3. Reeves, Dl, Bleimberg, J., Spector, J. Validation ANAM Battery in Multicenter Head Injury Rehabilitation. *Archives of Clinical Neuropsychology (1993) 8 (3): 262.*
4. Reeves DL, Schlegal R, Gilliland K, & Crabtree M,. UTC-PAB and the NATO/AGARD STRES Battery: results from standardization studies. Proceedings from the *1991 Medical Defense Bioscience review, Aberdeen Proving Grounds, MD, 1991*.
5. Thorne D, Genser S, Sing H, & Hegge F,. The Walter Reed Performance Assessment Battery. *Neurobehavioral Toxicology and Teratology (1985) 7: 415-418.*