NASA-STD-3001 Technical Brief

# Cognitive Workload Assessment Methods

## OCHMO-TB-032
## Rev C

## Relevant Technical Requirements

**NASA-STD-3001 Volume 1, Rev C**
[V1 3003] In-Mission Preventive Health Care
[V1 4011] Mission Cognitive Status
[V1 4014] In-Mission Completion of Critical Tasks
[V1 5002] Crewmember Training

**NASA-STD-3001 Volume 2, Rev D**
[V2 3006] Human-Centered Task Analysis
[V2 3101] Iterative Developmental Testing
[V2 10001] Usability
[V2 10002] Design-Induced Error
[V2 10003] Operability
[V2 10004] Controllability and Maneuverability During Manual Control (Handling Qualities – Level 1)
[V2 5052] Controllability and Maneuverability During Manual Control with Deficiencies (Handling Qualities – Level 2)
[V2 5006] Situation Awareness
[V2 5007] Cognitive Workload
[V2 10200] Physical Workload
[V2 10161] Automation System Status Provision

## Executive Summary

Cognitive Workload is the user's perceived level of mental effort that is influenced by many factors, particularly task load and task design.

The workload measurement enables standardized assessment of whether temporal, spatial, cognitive, perceptual, and physical aspects of tasks and the crew interfaces for these tasks are designed and implemented to support each other. Application of workload measurement for crew interface and task designs in conjunction with other performance measures, such as usability and design-induced error rates, helps assure safe, successful, and efficient system operations by the crew.

Designers need to consider the workload of the user when designing and producing an interface or designing a task. Low workload levels are associated with boredom and decreased attention to task, whereas high workload levels are associated with increased error rates and the narrowing of attention to the possible detriment of other information or tasks (Sandor et al., 2013). Humans perform best when they are neither bored nor overburdened, and when periods of work and rest are equitably mixed together.

Workload assessments should be integrated early and often through the engineering design life cycle so that related design decisions can be made from a data-driven perspective and ensure crew safety and performance.

# Background

- Previous studies have revealed the deleterious effects of workload that is either too high or too low (Measuring and Evaluating Workload: A Primer, NASA/TM—2010-216395).
    - High workload leads to humans hurrying performance, committing more errors, yielding poor accuracy, becoming frustrated, uncomfortable, and fatigued, and having poor awareness of their surroundings.
    - Low workload has been linked to high error rates, frustration, fatigue, and poor awareness of surroundings as they become bored, as their attention drifts, and as complacency sets in.

- Anecdotal reports from astronauts (Scheuring et al., 2007) indicate that at times of high intensity, workload can result in mental and physical fatigue.

- **Skylab 4:** Crewmembers reported that they quickly ran into difficulty due to work overload. The fast-paced schedule and workload of the mission initially caused crewmembers to consistently "feel" behind on tasks as well as demoralized. Mission Control personnel later acknowledged that the schedule had been such that it had not given the crewmembers adequate time in which to adjust to their environment (Uri, J. 2020).

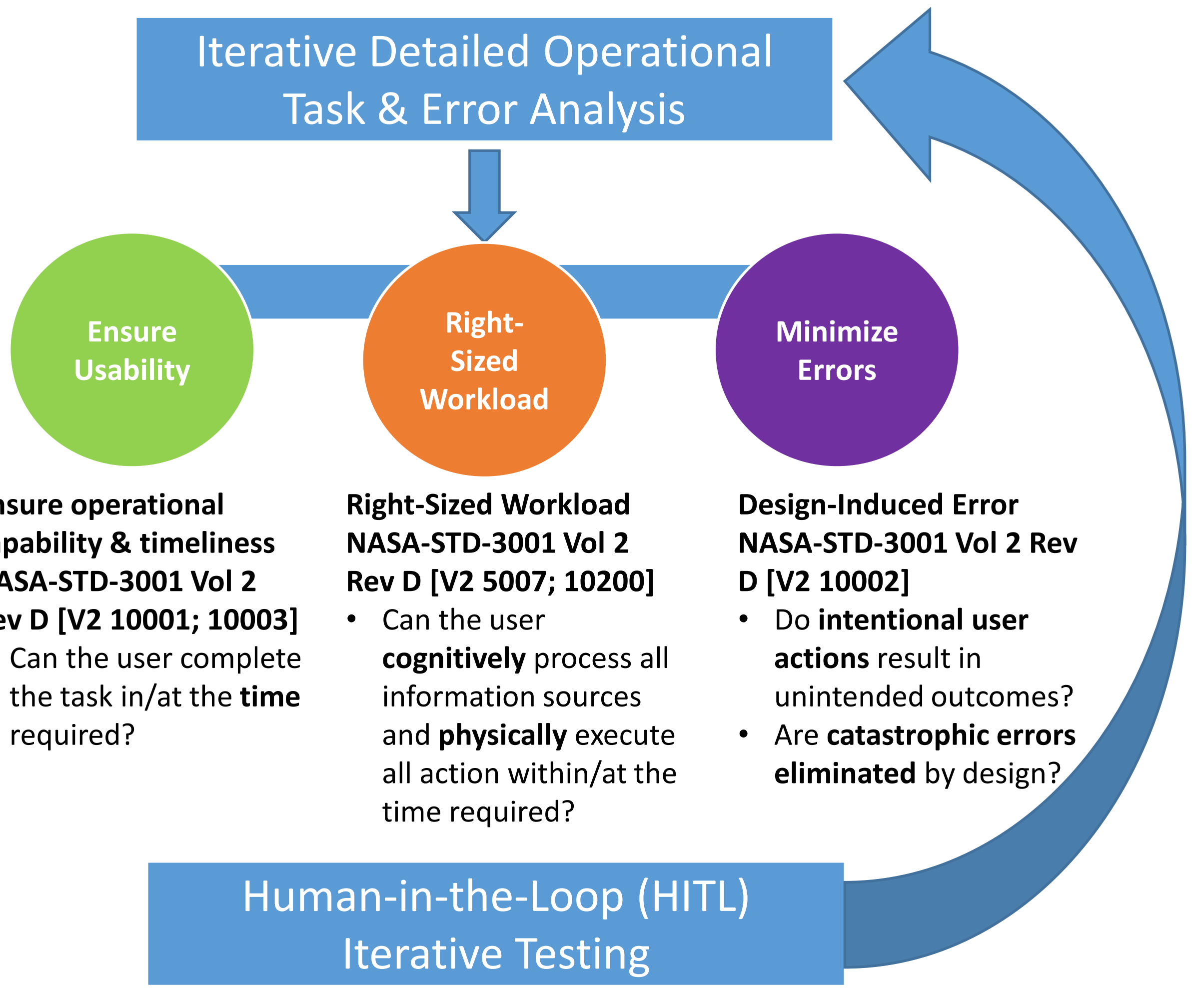
- **Apollo:** Some of the Apollo crews reported serious mental fatigue while they were performing lunar surface extravehicular activities (Scheuring et al., 2007).

- **ISS:** Astronauts experienced high tempo operations and high workloads (Behavioral Issues Associated with Long Duration Space Expeditions: Review and Analysis of Astronaut Journals Experiment 01-E104 (Journals): Final Report , 2010):
    - *"The fatigue was evident when a couple of minor mistakes were made today on some payload activities. The ground caught the mistake and helped me out. But it is an obvious indicator of fatigue. I feel that the workload is going up; these last few weeks seem to have been pretty taxing. I'm very tired."*

- **Titan 11 missile accident, 1980, Damascus, Arkansas:** Missile exploded, ejecting 9-megaton nuclear warhead. A series of human factors issues were reported in this incident. Crew had been working for 12 hours prior to the incident.

## Background

- Workload is closely linked to other human factors concepts such as usability, physical workload, anthropometrics, error rates, and handling qualities**.**
- Significant usability or handling quality issues will often drive high workload ratings.




**Relevant Technical Requirements**
[V2 5007] Cognitive Workload, [V2 10001] Usability, [V2 10002] Design-Induced Error, [V2 10003] Operability, [V2 10200] Physical Workload
*From: NASA-STD-3001 Volume 2, Rev D*

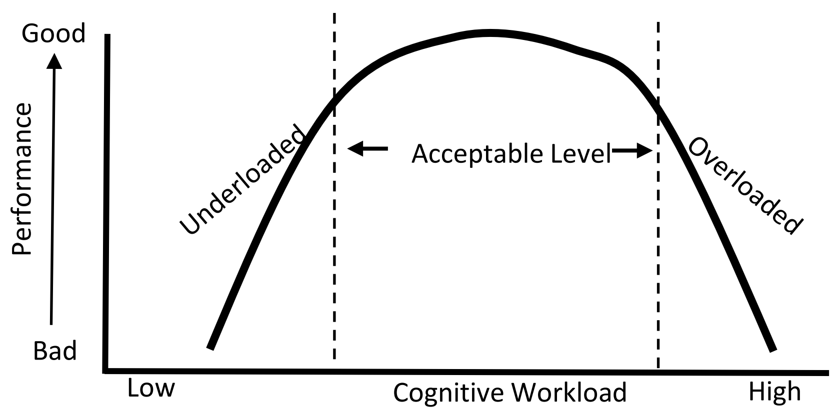
Reference NASA OCHMO-TB-005 Usability, Workload, & Error

# Background

Cognitive Workload can be suboptimal either because it is:

- **Low due to low arousal:** Technological advances (e.g., automation) introduced to improve system performance (and reduce workload) may sometimes create underload where the operator simply monitors the automation and removes them from the active control loop (Parasuraman & Riley, 1997); or may shift operator workload from one focus to another without achieving the expected reduction in workload. When an operator experiences excessively low task demands they may experience a state of mental underload.
- **High due to excessive task demands:** When operators are faced with excessive task demands and their attentional resources are exceeded, they become overloaded. Mental overload occurs when the demands of the task are so great that they are beyond the attentional capacity of the operator.

Overload and underload can be detrimental to task performance. Operators may become less likely to attend to potentially important sources of information and fail at the task.




**Workload Assessments**

Workload assessment tools such as the Bedford Scale or NASA Task Load Index (TLX) are available to help answer the following question**: "Can the user cognitively process all information sources and execute all actions within/at the time required**?"

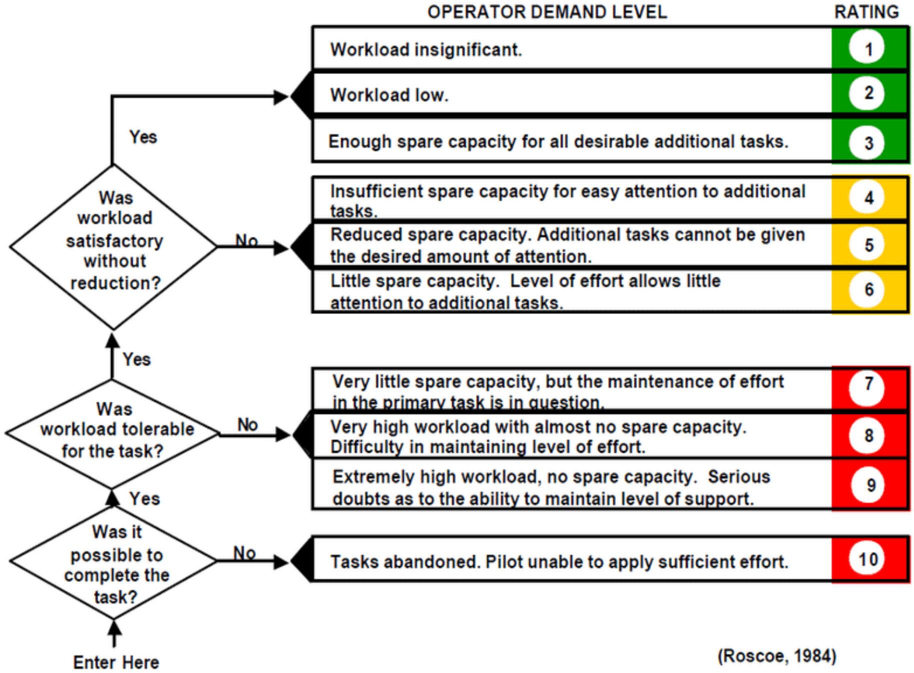
- NASA has determined that for space flight programs, workload should be assessed using the Bedford scale. Other types of scales (e.g., NASA Task Load Index (TLX)) are diagnostic or multi-dimensional, meaning that they allow the source of the workload to be localized. These types of scales are advantageous to use during the design phase, where modifications based on workload evaluations are possible. However, during the verification phase, the Bedford scale is the most appropriate.
- Metrics of Cognitive Workload measure the mental demands required of a person to perform a given task. Appropriate workload levels keep the crewmember engaged in the task, while allowing spare mental capacity to deal with concurrent tasks or issues.
- Some of the most safety-critical decisions and actions associated with operating a spacecraft are carried out in situations where the crew is multi-tasking, processing numerous inputs, and making decisions concerning multiple, possibly unrelated, problems. Work may also demand abrupt shifts between tasks performed alone and tasks relying on others' inputs.

# Reference Data

**The Bedford Scale**

The Bedford scale is a uni-dimensional rating scale designed to identify operator's spare mental capacity while completing a task. The single dimension is assessed using a hierarchical decision tree that guides the operator through a ten-point rating scale, each point of which is accompanied by a descriptor of the associated level of workload.

1. The operator rates their workload on a uni-dimensional scale from 1 to 10.
2. The operator is asked on a three-rank ordinal structure whether 1) it was possible to complete the task, (2) the workload was tolerable, or (3) the workload was satisfactory without reduction. Then, the operator ranks their workload on the respective rating scale end points (1-10) from workload insignificant to task abandoned.
3. This measures whether it was possible to complete the task, if workload was tolerable for the task, and if workload was satisfactory without reduction (Miller, 2001).




(Roscoe, 1984)

NASA-STD-3001, Volume 2 Rev D sets limits for Cognitive Workload using the Bedford scale:
**[V2 5007] Cognitive Workload** The system shall provide crew interfaces that result in Bedford Workload Scale ratings of 3 or less for nominal tasks and 6 or less for off-nominal tasks.

- Workload measurements for **nominal tasks (critical or frequent)** are limited to a Bedford Workload Scale rating of 3 or less.
- Workload measurements for **off-nominal tasks (non-critical or infrequent)** are limited to a Bedford Workload Scale Rating of 6 or less.
- Workload levels may be modulated (raised or lowered) through the combination of user-interface design and task design (e.g., task simplification, subtask combination and sequencing, and the distribution of tasks among multiple crewmembers and between crew and automation).

# Reference Data

**NASA Task Load Index (TLX)**
The NASA TLX may be preferred for developmental testing, due to its diagnostic properties. Other validated indicators of workload may be used by programs with approval from the Health and Medical Technical Authority.

The NASA TLX is a multi-dimensional rating scale for operators to report their mental workload. It uses six dimensions of workload to provide diagnostic information about the nature and relative contribution of each dimension in influencing overall operator workload. Operators rate the contribution made by each of the six dimensions of workload to identify the intensity of the perceived workload. The NASA TLX provides an overall score based on a weighted average of the six subscale ratings as described in the figure below.

- The degree to which each of the six factors contributes to the workload of the specific task to be evaluated from the rater's perspective is determined by their responses to pair-wise comparisons among the six factors.
- Magnitude ratings in each subscale are factors deemed most important in creating the workload of the task and are obtained after each performance of a task or task segment.

RATING SCALE DEFINITIONS

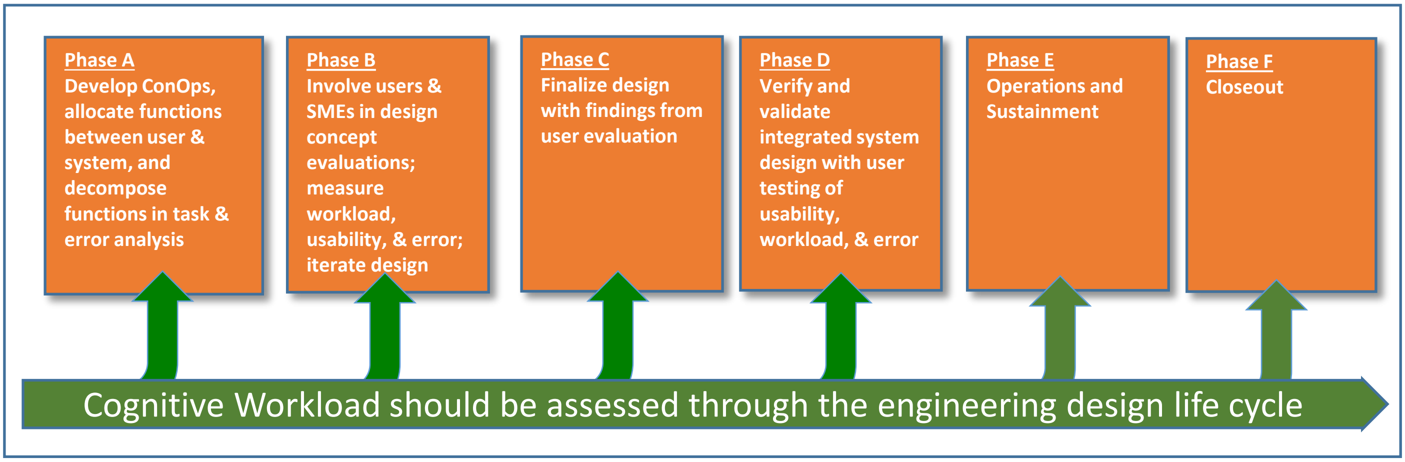
| Title | Endpoints | Descriptions |
|---|---|---|
| MENTAL DEMAND | *Low/High* | How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving? |
| PHYSICAL DEMAND | *Low/High* | How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious? |
| TEMPORAL DEMAND | *Low/High* | How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic? |
| EFFORT | *Low/High* | How hard did you have to work (mentally and physically) to accomplish your level of performance? |
| PERFORMANCE | *Good/Poor* | How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals? |
| FRUSTRATION LEVEL | *Low/High* | How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task? |

# Application

- Cognitive Workload is an important component of crew interaction with systems. Designers must consider it when designing hardware and software with crew interfaces, procedures, and operations.

- At the earliest stages of the design life cycle, integration of crew workload should focus on defining the various tasks that are relevant to workload.

- **Task analysis** is the method for identifying which crew and systems tasks will be performed during each mission phase, the hardware associated with the task, and whether the task is expected to contribute to crew workload. Task analysis should continue to mature as the design progresses. After task definitions, the next stage is to start assessing crew workload in a series of simulated vehicle tasks.




- When designing a human-system interface to support a crew task, designers are to assess the operation as part of a human-in-the-loop (HITL) simulation to determine the workload associated with that operation. If the Cognitive Workload is judged to be so high that a human has little or no spare capacity to deal with a concurrent problem, the task and supporting interfaces are to be redesigned.

- System design can aid in the selection and organization of relevant information to reduce the level of cognitive effort required to attend to a stimulus in the environment.

- Information should be prioritized so that the most important or critical information is displayed at all times and less important or critical information can be displayed upon a user's request (Federal Aviation Administration, 2011).

**[V2 3006] Human-Centered Task Analysis** Each human space flight program or project shall perform a human-centered task analysis to support systems and operations design. *From: NASA-STD-3001 Volume 2, Rev D*

# Application

- When users are required to monitor multiple displays, important events should occur in all of the displays to promote effective monitoring performance (Warm et al., 2008).

- Coding techniques that have strong attention-getting qualities (for example, color and flashing) should be used sparingly and judiciously (Federal Aviation Administration, 2011).

- Workload assessments should be integrated early and often through the engineering design life cycle so that related design decisions can be made from a data-driven perspective and ensure crew safety and efficiency. It is easier and more cost-effective to correct deficiencies in hardware or procedures that produce high crew workload during the early design phases rather than just before vehicle certification.

- The Bedford Workload Scale has been selected by NASA as the workload **verification** method for a number of program workload requirements. However, the NASA TLX may be preferred for developmental testing, due to its diagnostic properties. Other validated indicators of workload may be used by programs with approval from the Health and Medical Technical Authority.

- When using the Bedford scale assessment to evaluate Cognitive Workload, some guidelines need to be followed:

  - ✓ Participants must be briefed on what defines workload and should understand that they will be asked to rate their own workload. There are no “right” or “wrong” answers.

  - ✓ Operator comments should be solicited as part of the testing. The comments should be as specific as possible to enable the engineers to make appropriate changes to the design.

  - ✓ An appropriate number of operators should be tested to ensure meaningful results.

  - ✓ Testing should include worst case conditions (including off-nominal conditions) to simulate the stress the operator will be under while performing the operation.

  - ✓ Testing should also consider the operators physiological condition during performance of the task. Physiological conditions that influence performance include, but are not limited to: core body temperature, humidity levels, hydration levels, strength, aerobic capacity, and orthostatic intolerance.

# Back-Up

## Major Changes Between Revisions

Rev B → Rev C

- Updated information to be consistent with NASA-STD-3001 Volume 1 Rev C and Volume 2 Rev D.

Rev A → Rev B

- Added additional information and application notes on specific technical requirements.

Original→ Rev A

- Updated information to be consistent with NASA-STD-3001 Volume 1 Rev B and Volume 2 Rev C.

View the current versions of NASA-STD-3001 Volume 1 & Volume 2 on the [OCHMO Standards website]

# Referenced Technical Requirements

**NASA-STD-3001 Volume 1 Revision C**

**[V1 3003] In-Mission Preventive Health Care** All programs shall provide training, in-mission capabilities, and resources to monitor physiological and psychosocial well-being and enable delivery of in-mission preventive health care, based on epidemiological evidence-based probabilistic risk assessment (PRA), individual crewmember needs, clinical practice guidelines, flight surgeon expertise, historical review, mission parameters, and vehicle derived limitations. These analyses consider the needs and limitations of each specific vehicle and design reference mission (DRM) with particular attention to parameters such as mission duration, expected return time to Earth, mission route and destination, expected radiation profile, concept of operations, and more. In-mission preventive care includes, but is not limited to: (see NASA-STD-3001 Volume 1 Rev C for full technical requirement).

**[V1 4011] Mission Cognitive Status** Pre-mission, in-mission, and post-mission crew behavioral health and crewmember cognitive state shall be within clinically accepted values as judged by behavioral health evaluation.

**[V1 4014] Completion of Critical Tasks** The planned number of hours for completion of critical tasks and events, workday, and planned sleep period shall have established limits to assure continued crew health and safety.

**[V1 5002] Crewmember Training** Beginning with the astronaut candidate year, general medical training, including, but not limited to, first aid, cardiopulmonary resuscitation (CPR), altitude physiological training, carbon dioxide exposure training, familiarization with medical issues, procedures of space flight, psychological training, and supervised physical conditioning training shall be provided to the astronaut corps.

**NASA-STD-3001 Volume 2 Revision D**

**[V2 3006] Human-Centered Task Analysis** Each human space flight program or project shall perform a human-centered task analysis to support systems and operations design.

**[V2 3101] Iterative Developmental Testing** Each human space flight program or project shall perform iterative human-in-the-loop (HITL) testing throughout the design and development cycle.

**[V2 10001] Usability** The system shall provide crew interfaces that result in a NASA-modified System Usability Scale (NMSUS) score of 85 or higher.

**[V2 10002] Design-Induced Error** The system shall provide crew interfaces that result in the maximum observed error rates listed in Table 5.1-1—Maximum Observed Design-Induced Error Rates.

**[V2 10003] Operability** The system shall provide interfaces that enable tasks to be performed successfully within the appropriate time limit and degree of accuracy.

**[V2 10004] Controllability and Maneuverability During Manual Control (Handling Qualities – Level 1)** The spacecraft shall exhibit Level 1 handling qualities (Handling Qualities Rating (HQR) 1, 2 and 3), as defined by the Cooper-Harper Rating Scale, during manual control of the spacecraft's flight path and attitude when manual control is the primary control mode or automated control is nonoperational.

**[V2 5052] Controllability and Maneuverability During Manual Control with Deficiencies (Handling Qualities – Level 2)** The system shall exhibit Level 2 (HQR 4-6) or better handling qualities during manual control in all other scenarios not specified in [V2 10004] Handling Qualities – Level 1.

# Referenced Technical Requirements

View the current versions of NASA-STD-3001 Volume 1 & Volume 2 on the OCHMO Standards website

**NASA-STD-3001 Volume 2 Revision D**

**[V2 5006] Situation Awareness** Systems shall provide the Situation Awareness (SA) necessary for efficient and effective task performance and provide the means to recover SA, if lost, for anticipated levels of crewmember capability and anticipated levels of task demands.

**[V2 5007] Cognitive Workload** The system shall provide crew interfaces that result in Bedford Workload Scale ratings of 3 or less for nominal tasks and 6 or less for off-nominal tasks.

**[V2 10200] Physical Workload** The system shall provide crew interfaces that result in a Borg-CR10 rating of perceived exertion (RPE) of 4 (somewhat strong) or less.

**[V2 10161] Automation System Status Provision** The automated system shall provide the human operator with the following information:

a. system state (e.g., position, location, hazardous condition, running, paused, faulted, completed, overridden, stopped, readiness)

b. projection of future state, including failure or decrements in performance (e.g., battery power versus traverse distance and assessment of uncertainty in projection of future state) and mode (e.g., Full/Partial/Manual/Test)

c. system health

d. configuration information (e.g., setup/input parameters, initial conditions, and terminating conditions)

# Reference List


1. Basner M, Dinges DF, Mollicone DJ, Savelev I, Ecker AJ, et al. (2014) Psychological and Behavioral Changes during Confinement in a 520-Day Simulated Interplanetary Mission to Mars. PLOS ONE 9(3): e93298. https://doi.org/10.1371/journal.pone.0093298
2. Behavioral Issues Associated with Long Duration Space Expeditions: Review and Analysis of Astronaut Journals Experiment 01-E104 (Journals): Final Report. (2010). https://lsda.jsc.nasa.gov/lsda_data/dataset_inv_data/ILSRA_2001_104__1740256372_.pdf_Expedition_8_ILSRA-2001-104_2011_31_010100.pdf
3. Fatigue Risk Management in Aviation Maintenance: Current Best Practices and Potential Future Countermeasures, Federal Aviation Administration, June 2011
4. Flight Test Assessments of Pilot Workload, System Usability, and Situation Awareness of TASAR,HFES International Annual Meeting, Human Factors Principal Investigator, NASA Langley Research Center, 2016
5. Measuring and Evaluating Workload: A Primer. Stephen M. Casner, NASA Ames Research Center, Moffett Field: https://pdfs.semanticscholar.org/3156/ae866b00a2ec9495a5b8b3df5422196d179a.pdf?_ga=2.172460323.140449752.1579710142-474163027.1579710142
6. Risk of Performance Decrements and Adverse Health Outcomes Resulting from Sleep Loss, Circadian Desynchronization, and Work Overload Rev. 2,HSRB-CR-SA-01743, 2019
7. Sandor, A., Schuh, S.V., and Gore, B.F. (2013). Evidence Report: Risk of Adverse Cognitive or Behavioral Conditions and Psychiatric Disorders. *Human Research Program*, National Aeronautics and Space Administration Lyndon B. Johnson Space Center Houston, Texas. https://humanresearchroadmap.nasa.gov/Evidence/reports/TASK.pdf
8. Psychological and Behavioral Changes during Confinement in a 520-Day Simulated Interplanetary Mission to Mars, March 2014: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0093298
9. Stress and Cognitive Workload: https://www.nap.edu/read/5436/chapter/8#152
10. The Use Of Pilot Rating In The Evaluation Of Aircraft Handling Qualities by George E. Cooper ,April 1969,NASA TN D-5153.
11. Miller, S. (2001). Workload Measures Technical Report, N01-006. Iowa City, IA, USA. University of Iowa. http://www.nads-sc.uiowa.edu/publicationStorage/200501251347060.N01-006.pdf
12. Scheuring, R.A., et al. (2007). The Apollo Medical Operations Project: Recommendations to Improve Crew Health and Performance for Future Exploration Missions and Lunar Surface Operations. NASA/TM–2007–214755. https://www.nasa.gov/wp-content/uploads/2015/05/medical_distributed_planning_systems_lessons_learned.pdf
13. Uri, J. (2020). The Real Story of the Skylab 4 “Strike” in Space. *NASA*. https://www.nasa.gov/history/the-real-story-of-the-skylab-4-strike-in-space/
14. Parasuraman, R., and Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors, 39*(2): 230-253.
15. Warm, J.S., Parasuraman, R., and Matthews, G. (2008). Vigilance Requires Hard Mental Work and Is Stressful. *Human Factors, 50*(3): 433-441.