Executive Summary

Cognitive Workload, like usability, is a complex construct but essentially means the amount of effort people have to exert mentally to use the interface.

Designers need to consider the workload of the user when designing and producing an interface or designing a task. Low workload levels have been associated with boredom and decreased attention to task, whereas high workload levels have been associated with increased error rates and the narrowing of attention to the possible detriment of other information or tasks (Sheridan, 2002). Humans do their 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.

Summary of Standards

NASA-STD-3001 Vol. 1
4.4.3.5.2 General Health and Well Being

NASA-STD-3001 Vol. 2
[V2 3006] Human-Centered Task Analysis
[V2 5004] Cognitive Capabilities
[V2 5005] Time and Performance
[V2 5007] Nominal Cognitive Workload
[V2 5008] Off-Nominal Cognitive Workload
[V2 10001] Usability Acceptance Criteria
[V2 10003] Crew Interface Efficiency

Space shuttle Challenger
Background

• The earliest studies quickly revealed the deleterious effects of workload that is either too high or too low (Measuring and Evaluating Workload: A Primer, NASA/TM—2010-216395).

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

• Skylab 3: Crewmembers reported that they quickly ran into difficulty due to work overload. The fast-paced schedule and workload of the mission had initially caused these crew members to consistently “feel” behind on tasks as well as demoralized. At the start of the 45th day of their 59-day mission, the crewmembers elected to have a sit-down, during which they refused to perform scheduled tasks. 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 (Cooper Jr., 1996).

• Apollo: Some of the Apollo crews reported serious mental fatigue while they were performing lunar EVAs (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.”
Background

• NASA STD-3001 Volume 2 defines cognitive workload (right sized workload in diagram below) as the number of mental operations or activities that are conducted simultaneously or in close succession. Physical workload refers to the number of individual physical activities that are conducted simultaneously or in close succession. This technical brief focuses on cognitive workload.

• Workload is closely linked with other human factors concepts such as usability, physical workload, anthropometrics, error rates and handling qualities.

• Significant usability or handling qualities issues will often drive high workload ratings.

Relationship of usability, workload and 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, 1987). Or may shift operator workload from one locus 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.
Background

Workload Assessments

Workload assessment tools such as 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 action within/at time required?”

• In space flight, the primary concern is avoiding unnecessarily high workload levels, given that space flight is generally a high-stress environment. Therefore, the process described below focuses on measuring workload with the goal of keeping workload at a level that does not negatively affect performance.

• NASA has determined that for space flight programs workload should be assessed using the Bedford scale. Compared to other scales, the Bedford determines whether workload is tolerable for the task.

• Many 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.

- The operator has to rate his/her workload on a uni-dimensional rating scale from 1 to 10.
- The operator is asked on a three-rank ordinal structure wether 1) it was possible to complete the task, (2) the workload was tolerable, or (3) the workload was satisfactory without reduction and after this first decision the operator has to rank his workload on the respective rating scale end points (1-10) from workload insignificant to task abandoned.
- It ranks whether it was possible to complete the task, if workload was tolerable for the task, and if workload was satisfactory without reduction (Rehmann, 1995).

![Bedford Workload Scale Diagram](image)

NASA STD - 3001, Volume 2 set limits for cognitive workload using the Bedford scale

- Workload measurements for **nominal tasks**: Excessive workload demands on any one task can cause the operators to exclusively focus on one problem or approach to performing the tasks, leaving little or no spare capacity to deal with any other problems that may occur.
  - [V2 5007] The system shall provide crew interfaces that, when used to perform nominal crew tasks, result in **Bedford Workload Scale ratings of 3 or less** (or equivalent rating on another validated workload scale).

- Workload measurements for **off-nominal tasks**: During off-nominal events, it is important to ensure crewmembers’ workload is not excessively high. High workload can leave an operator with little or no spare mental capacity to contend with additional demands that, during an off-nominal situation, could have significant negative outcomes.
  - [V2 5008] The system shall provide crew interfaces that when used to perform off-nominal crew tasks result in a **Bedford Workload Scale rating of 6 or less** (or equivalent rating on another validated workload scale).
Reference Data

NASA TLX

The NASA Task Load Index (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.

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 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 ratings the six subscales 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 obtained after each performance of a task or task segment. Ratings of factors deemed most important in creating the workload of the task.

<table>
<thead>
<tr>
<th>Rating Scale Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
</tr>
<tr>
<td>MENTAL DEMAND</td>
</tr>
<tr>
<td>PHYSICAL DEMAND</td>
</tr>
<tr>
<td>TEMPORAL DEMAND</td>
</tr>
<tr>
<td>EFFORT</td>
</tr>
<tr>
<td>PERFORMANCE</td>
</tr>
<tr>
<td>FRUSTRATION LEVEL</td>
</tr>
</tbody>
</table>
Application Notes

• 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 would be 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 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 all the time and less important or critical information can be displayed upon a user’s request (Avery & Bowser, 1992).
Application Notes

- 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., 1996).

- Coding techniques that have strong attention-getting qualities (for example, color and flashing) should be used sparingly and judiciously (National Air Traffic Services, 1999).

- 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 Task Load Index (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 the Cognitive workload, some guidelines need to be followed:
  
  ✓ Participants need to 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 a part of the testing. The comments should be as specific as possible to enable the engineers to make appropriate changes to the design.

  ✓ 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/not limited to: core body temperature, humidity levels, hydration levels, strength, aerobic capacity, and orthostatic intolerance etc.
References

- Fatigue Risk Management in Aviation Maintenance: Current Best Practices and Potential Future Countermeasures, Federal Aviation Administration, June 2011
- NASA/SP-2010-3407 Human Integration Design Handbook (HIDH)
- NASA/TP-2014-218556 Human Integration Design Process
- Measuring and Evaluating Workload: A Primer Stephen M. Casner, Ph.D, NASA Ames Research Center, Moffett Field: https://pdfs.semanticscholar.org/3156/ae866b00a2ec9495a5b8b3df5422196d179a.pdf?_ga=2.172460323.140449752.1579710142-474163027.1579710142
- Risk of Performance Decrement and Adverse Health Outcomes Resulting from Sleep Loss, Circadian Desynchronization, and Work Overload, CR-HSRB-14-014-R1, 2015
- Risk of Performance Decrement and Adverse Health Outcomes Resulting from Sleep Loss, Circadian Desynchronization, and Work Overload Rev. 2, HSRB-CR-SA-01743, 2019
- Risk of Adverse Cognitive or Behavioral Conditions and Psychiatric Disorders, Evidence Report, Human Research Program, National Aeronautics and Space Administration Lyndon B. Johnson Space Center Houston, Texas
- 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
- Stress and cognitive workload :https://www.nap.edu/read/5436/chapter/8#152
- The Use Of Pilot Rating In The Evaluation Of Aircraft Handling Qualities by George E. Cooper, April 1969, NASA TN D-5153.
Back-Up
Referenced Standards

Volume 1

4.4.3.5.2 Countermeasures shall be provided to address issues of human factors and general crew health and well-being, including considerations for hygiene, privacy, nutrition, crew schedule, workload, Earth observation, and leisure activities.

Volume 2

[V2 3006] Each human spaceflight program or project shall perform a task analysis to support hardware and operations design.

[Rationale: A detailed task analysis of crew required activities is required to determine appropriate human spacecraft designs and layout along with the appropriate operational activities. This task analysis is utilized by numerous other standards such as net habitable volume, cognitive workload, situational awareness, display design, information management, EVA suit mobility, etc.]

[V2 5004] Cognitive performance capabilities shall be accommodated in the design of all system elements that interface with the crew for all anticipated levels of crew capability and all anticipated levels of task demands.

[Rationale: Accommodating cognitive performance capabilities is important to ensure optimal task performance and crew safety. Design of hardware, including displays and controls, are to take into account the capabilities and limitations of humans to acquire, interpret, and retain information such that the relevant information is available and intelligible. This is especially important during spaceflight, where microgravity can cause deconditioning and affect spatial orientation, radiation can induce acute cognitive deficits, and where stress can affect several cognitive processes. For detailed discussions regarding the effects of stress on cognitive performance, see chapter 5, Human Performance Capabilities, of the HIDH. Determination of anticipated levels of crew capability and anticipated levels of task demands is based on a detailed task analysis with consideration of all nominal, off-nominal, and emergency scenarios, including those of low probability.]

[V2 5005] The ability to perform tasks in a timely and accurate manner shall be accommodated in the design of all system elements that interface with the crew for anticipated levels of crew capability and anticipated levels of task demands.

[Rationale: Factors to be considered in design include, but are not limited to: time to complete the task; time to train for the task; consequence and time to recover from errors; the nature, type (e.g., independent versus team tasks) and environmental conditions of the task; and the state of the human and the team (e.g., deconditioned, lack of adaptation to microgravity, sleep deprived). Some prominent aerospace accidents have been traced to the human’s inability to perform emergency operations under the time demands and environmental challenges present, emphasizing the need for design that integrates crew capability with task demands. Determination of anticipated levels of crew capability and anticipated levels of task demands is based on a detailed task analysis.]
[V2 5007] The system shall provide crew interfaces that, when used to perform nominal crew tasks, result in Bedford Workload Scale ratings of 3 or less (or equivalent rating on another validated workload scale).

[Rationale: 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. Likewise, environmental stressors such as radiation and altered atmospheric composition or pressure may impede the ability to changes in cognitive workload. Excessive workload demands on any one task can cause the operators to exclusively focus on one problem or approach to performing the tasks, leaving little or no spare capacity to deal with any other problems that may occur. Therefore, having designed a human-system interface to support a crew task, designers are to assess the operation as part of a human-in-the-loop 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. The Bedford Workload Scale has been selected by NASA as the workload verification method for a number of program workload requirements, as of the publication of this volume. However, the NASA Task Load Index (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.]

[V2 5008] The system shall provide crew interfaces that when used to perform off-nominal crew tasks result in a Bedford Workload Scale rating of 6 or less (or equivalent rating on another validated workload scale).

[Rationale: During off-nominal events, it is important to ensure crewmembers’ workload is not excessively high. High workload can leave an operator with little or no spare mental capacity to contend with additional demands that, during an off-nominal situation, could have significant negative outcomes. The Bedford Workload Scale has been selected by NASA as the workload verification method for a number of program workload requirements, as of the publication of this volume. However, the 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.]
[V2 10001] Each program shall define usability acceptance criteria for crew interfaces.

[Rationale: Usable crew interfaces allow users to achieve task goals efficiently, effectively, and with satisfaction (ISO-9241-11). Efficiency, effectiveness, and satisfaction are the three major components of usability; therefore, acceptance criteria for all three should be defined by every program. Usability testing must be part of the verification process. Effective crew interfaces allow users to achieve specified tasks with accuracy and completeness. Efficient crew interfaces allow users to expend appropriate amounts of resources, e.g., time, workload, to achieve the effectiveness necessary in a specified context of use. Users are satisfied with a crew interface if they are willing to use a crew interface and have positive subjective responses and attitudes toward the crew interface.]

[V2 10003] The system shall provide crew interfaces that are efficient, e.g., with reduced training time, task time, and frustration.

[Rationale: Efficiency is determined by time on task, number of errors made, and the training demands. Maximizing efficiency should not be achieved through high training demands or performance pressure. Crew interfaces that do not maximize efficiency can result in frustration.]