Human Systems
integration division

Workload and Stress, and the Mission Centric Human Performance Measurement Interest Groups

HFE DoD TAG Wednesday, October 27, 2010

Brian F. Gore, PhD
SJSURF/NASA Ames Research Center
Human Systems Integration Division
Moffett Field, CA

Human Systems
integration division

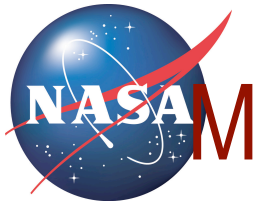
Speakers

- Brian Gore, PhD – SJSURF/NASA ARC
  - "Workload scales, measurement and management: Considerations for human-system integration in extreme environments"
- Lowell Staveland, MS – SHFE, Inc.
  - "Recollections on developing NASA TLX"
- Mark Pestana
  - "A pilot's perspective on workload"
- Robert McCann, PhD - NASA ARC
  - "End-of-trial workload ratings and real-time operator behavior: Making the connection"
- Rahel Rudd, PhD – WPAFB – Government employees only
  - "Workload and stress"

Workload Scales, Measurement and Management: Considerations for Human-System Integration in Extreme Environments

Brian F. Gore, Ph.D.
SJSURF/NASA Ames Research Center
Human Systems Integration Division
Moffett Field, CA

HFE DoD TAG, October 27, 2010



Motivation: Workload and Requirements Verification

- Designs need specifications but specifications can not constrain how designs are met*
- Requirements specified redline threshold for workload/ usability/handling design criteria*
  - HSIR Rev A specified Bedford scale to be used for all crew tasks
  - HSIR Rev B removed specific requirement in favor of a general requirement for workload and usability
  - HSIR Rev C modified values to close the “to-be-resolved” design values on the TLX scale, reduce the number of requirements, and change the scope of the usability calculations
  - HSIR Rev D modified the criteria back to Bedford given SME input

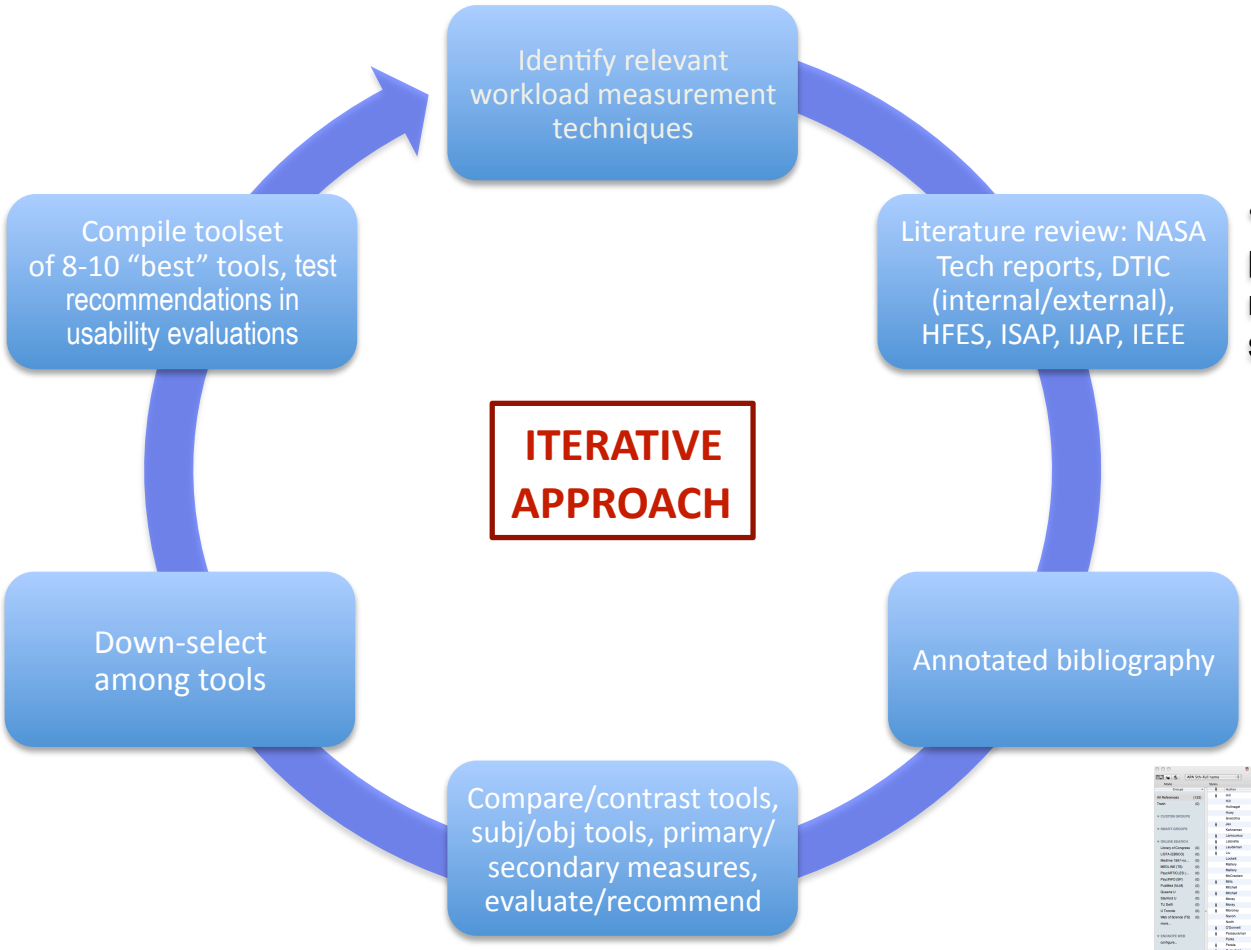
*Rochlis, J. (2010). Workload requirements in practice. In Gore, B.F., Macramalla, S., & Salud, E. (2010). A workshop on workload scales, measurement, and management for long duration space operations (No. NASA/CP-2010-216398). Washington, D.C.: National Aeronautics and Space Administration.

Human Systems integration division

Approach

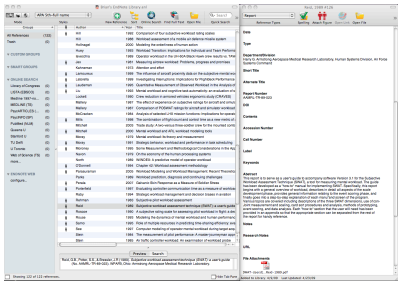
• 15+ years studying workload

• No one, single approach, workload primer helps, **Dose concept** proposed



• Principled approach, peer reviewed articles, relevant to the domain, start broad then refine.

• dimensionality, *thresholds*, operational environment, requirement and design-related approaches



State of the Art of Workload Measurement

Workload Defined

- A set of task demands, the effort to attain the task demands, or the accomplishment of the task demands*
- Workload arises from the interaction between a particular task and the performer
- May be considered physical or cognitive (mental)
  - Workload can be suboptimal either because it is too low due to low arousal, or too high due to excessive task demands, poor equipment design, or difficult environmental conditions
- An operator's perception of the workload

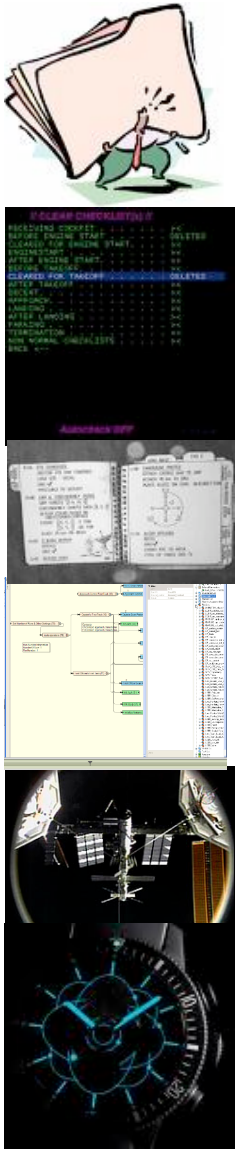




*Gartner, & Murphy, 1979; Gawron, 2008

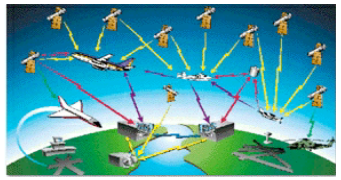
Workload Drivers

- Task structure
- Performance criteria and schedule
- Task schedule
- Rate of presentation / load
- Complexity of task
- Variability of task demands
- Task duration
- Time estimation
- Automation

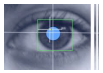


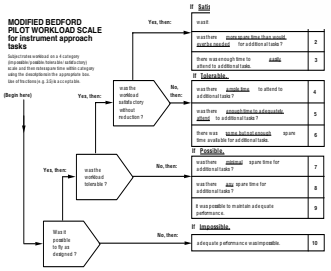
* Hart, 1978; Hart & Staveland, 1988; Casner, 2005; Hancock & Chignell, 1988; Roscoe, 1984; Stein, 1985; Reid, 1989; Gawron, 2008; NRC, 1993

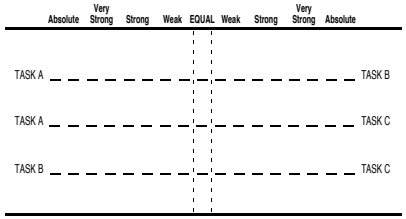
Workload Measures



- Workload measures and techniques
  - Designed for highly procedural aviation contexts
  - Typically applied to tasks *over short periods of time**
- Primary task measures (speed, accuracy, response profiles)
- Secondary task measures measuring residual attention or capacity
- Objective measures (Physiological)
- **Subjective measures** (Ratings)
- Issues
  - Intrusiveness of the measure
  - Context
  - Sensitivity
  - Reliability
  - Diagnosticity
  - Acceptability of relying on one primary measure or one secondary measure
  - Purpose of the workload measure
  - Affect, emotional states, and social factors
  - Workload threshold (notions of underload and overload)



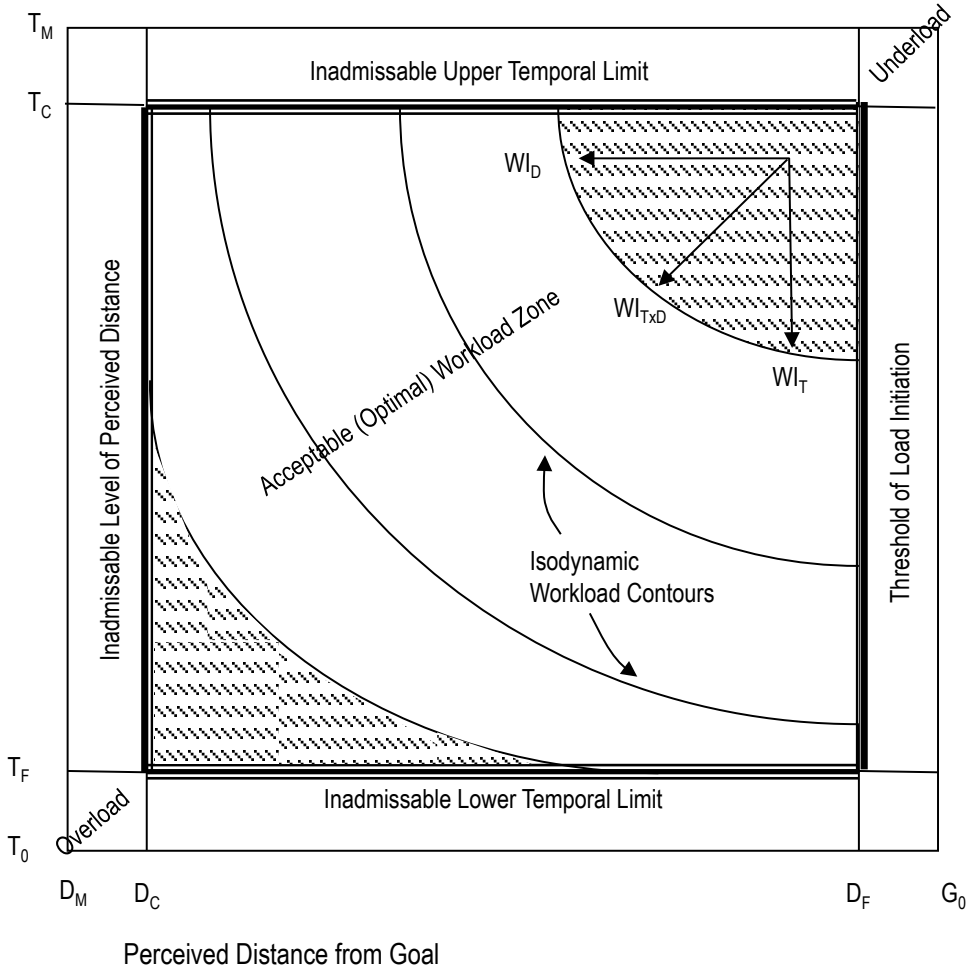




* Hart, 1978; Hart & Staveland, 1988; Casner, 2005; Hancock & Chignell, 1988; Roscoe, 1984; Stein, 1985; Reid, 1989; Gawron, 2008; NRC, 1993

Workload Thresholds

Effective
Time
For Action



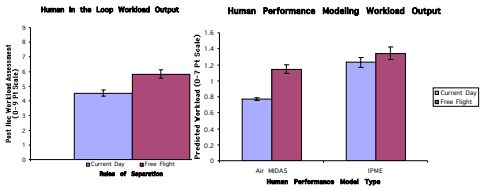
LEGEND
T_M - Conceptual Maximal Time for an activity that links a series of task related activities
T_C -Time horizon for coherent action
T_F - Floor for the operator response time
T_0 - Immediate Present
D_M - Amount of reconciliation of the perceived distance that must occur
D_C - Perceived maximal distance threshold (specific to an individual)
D_F - Discontinuity resulting from perceived distance from goal state
G_0 - Goal State
WI_D - Workload Increase due to Distance
WI_T - Workload Increase due to Time
WI_{DXT} - Workload Increase due to Time X Distance
Isodynamic Contours - points of equal workload loading
═ - Threshold of acceptable levels for human performance

But do these notions apply to space operations?

Adapted from Hancock & Chignell (1988)

Subjective Workload Measures

- Ask the human operator to describe the workload they experience when performing a task
- Does not attempt to measure anything about the task that the user is performing or their performance on the task
- Focuses entirely on the human operator's subjective impressions about their workload
- Two basic varieties have been developed:
  - *Subjective numerical measurement techniques* ask the human operator to assign a numerical or ordinal value to the workload that they are currently experiencing while working in a particular task situation
  - *Subjective comparative measurement techniques* ask the human operator to make comparisons between two or more tasks situations and say which situation results in the higher (or lower) workload



Casner & Gore, 2010

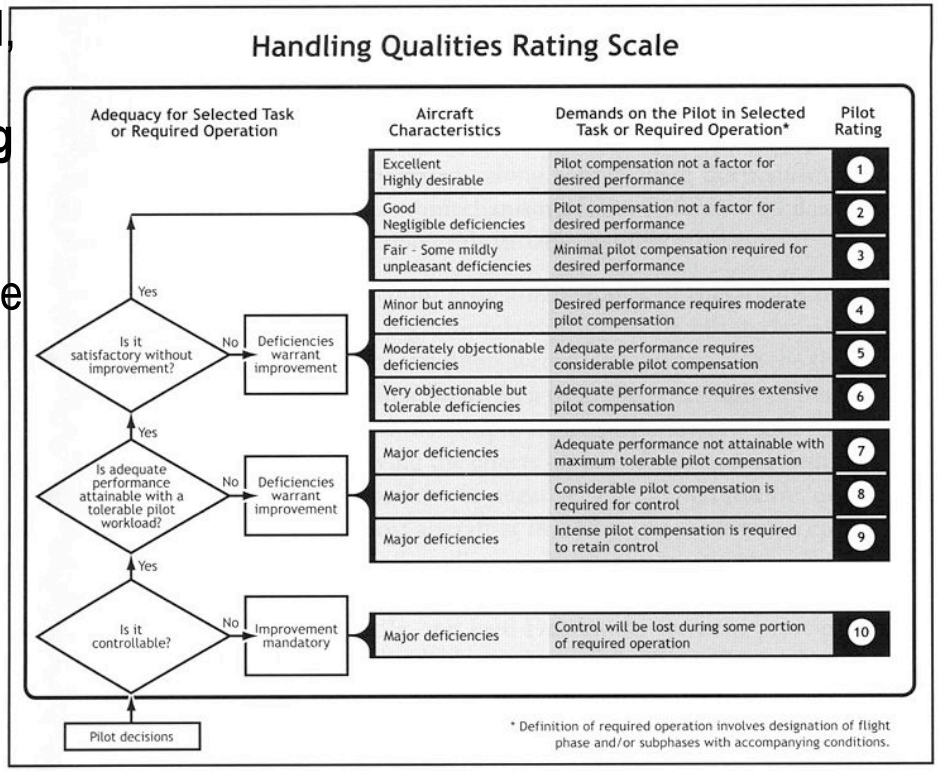
Human Systems integration division

Workload Measures Toolset

Scale	Strengths	Limitations	Dimensions	Anchors	Applicability
Cooper-Harper	Sensitive to taskload; reflects differences in performance	Ordinal scale, instruction set needs to be clear, task needs to be fully defined	Single	Yes	Aircraft handling, real world and empirical experiments
Modified Cooper-Harper	Valid Reliable estimate of overall workload	Not as sensitive as TLX; Requires memorable events in the simulation; Not an interval scale*	Single	Yes	Empirical experiments only, not real world evaluations
Bedford	Ease of use, anchors	Done independently from primary task Not an interval scale; Notion of "spare capacity" is ambiguous (it it time, mental capacity, physical capacity, etc?)	Single	Yes	Aerospace, nuclear, aviation, surface transportation, real world and empirical experiments
Instantaneous Self Assessment	Very simple to collect/easy to administer/understand (Castle & Leggatt, 2002), not intrusive – allows real time estimates of perceived workload Relative comparisons	Definition of workload can be misinterpreted; Response clusters around different portions of the scale, differences in workload perceptions is difficult to tease out; Relationship between workload and quality of performance	Single	Yes	Aerospace, nuclear, aviation, surface transportation, real world and empirical experiments
NASA TLX	Validated, multi-dimensional measure, less subject to memory errors, favored by subjects over SWAT, Cooper-Harper, high correlation with other workload measures, reliable, easy to use, portable.	No behavioral anchors; Paper test is subject to transcription error; Weightings add undue complexity	Multiple	No	Aerospace, nuclear, aviation, surface transportation, real world and empirical experiments

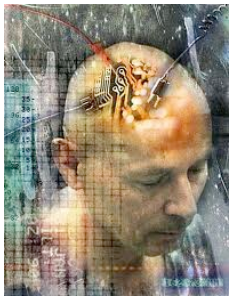
Subjective Measures: Cooper-Harper

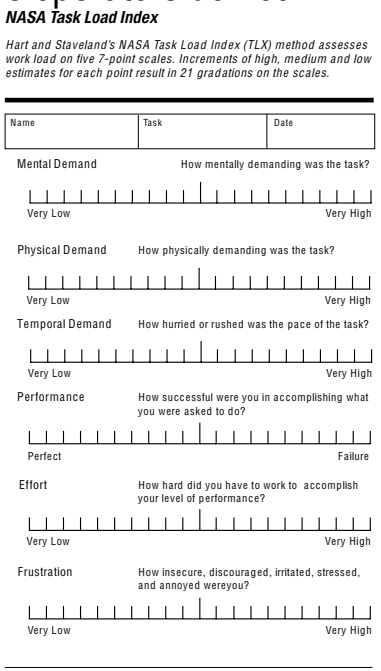
- One of the earliest rating scales developed, still used today
- Subjective evaluations of **aircraft handling qualities**
- Raters make a series of decisions, each of which discriminates between two or three alternatives
- Numeric rating range 1-10
- Intuitive, anchors
- Raters may read the scale each time or can rely on their memory
- Modified versions of the scale exist that retain the decision tree format, but substitute terms that address workload more directly
- Limitations: Ordinal scale, instruction needs to be clear, task needs to be fully defined, Uni-dimensional, poor diagnosticity & poor reliability



Subjective Measures: NASA Task Load Index (TLX)

- Subjective workload assessments on operator(s) working in human-machine systems, estimates along mental, physical, temporal, performance, frustration, effort sub-scales
- Advantages
  - Favored by subjects vs SWAT, Cooper-Harper
  - Correlated with other measures of workload; Reliable for repeated measures*
  - Easy to implement/use
  - Less subject to memory effects (workload is measured in situ)
  - More detailed analysis of the workload given its multidimensionality
  - Portable and can be used in operational experiments
  - Attempts to accommodate for biases in subject performance
- Disadvantages
  - Paper version post-processing time, transcription errors possible
  - No anchors
  - More time-consuming to administer than other techniques given the 6 sub-scales, rank ordering
  - Scale loading problems operators do not always think of a value of 50 as the middle and move linearly toward the two ends of the scale as perceived workload rises and falls



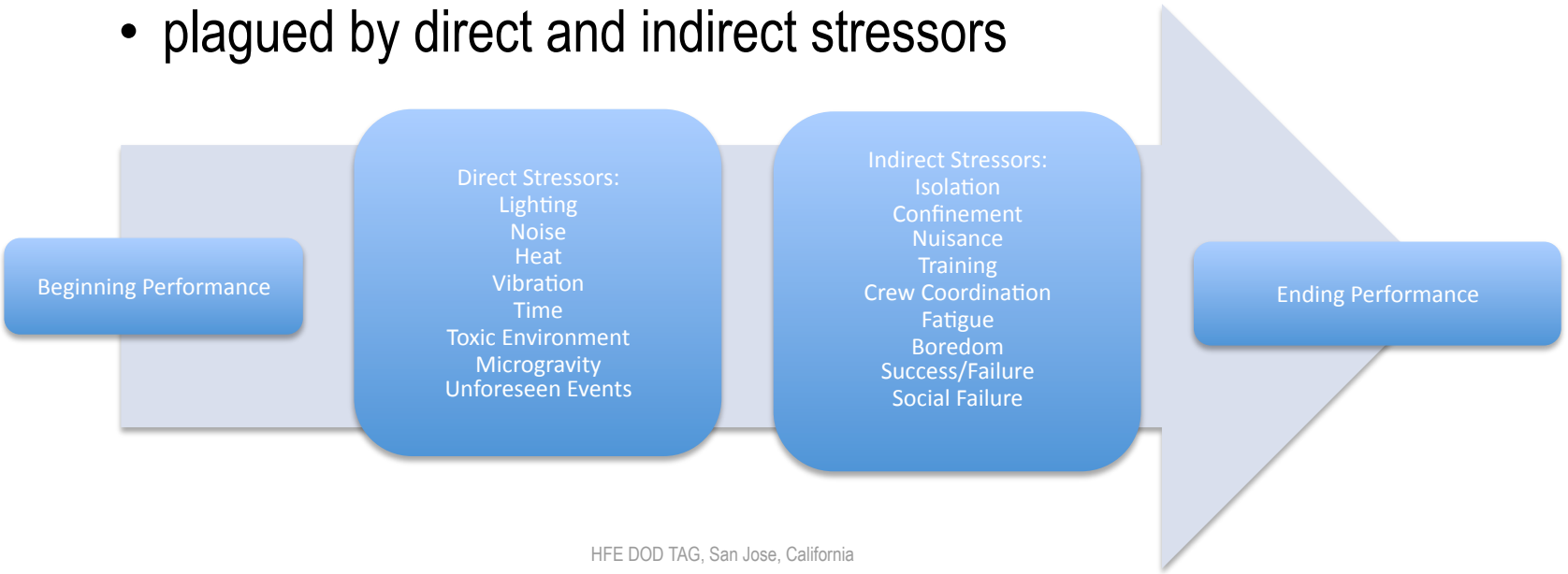


* Battiste & Bortolussi, 1988; Hill et al., 1992

The World of Space Operations

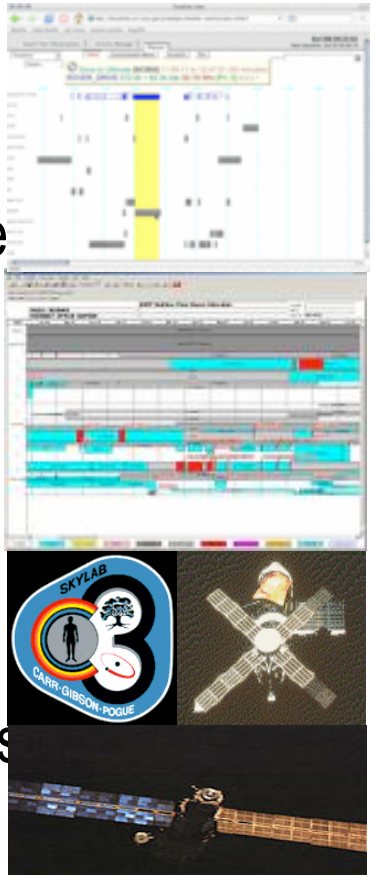
Scope: Long-duration Space Missions

- What are long-duration space missions?
  - highly procedural but highly repetitive tasks
  - extended missions lasting from weeks to 2.5 years
  - multiple crewmembers simultaneously sharing in the task performance
  - plagued by direct and indirect stressors



Space Mission Workload Challenge

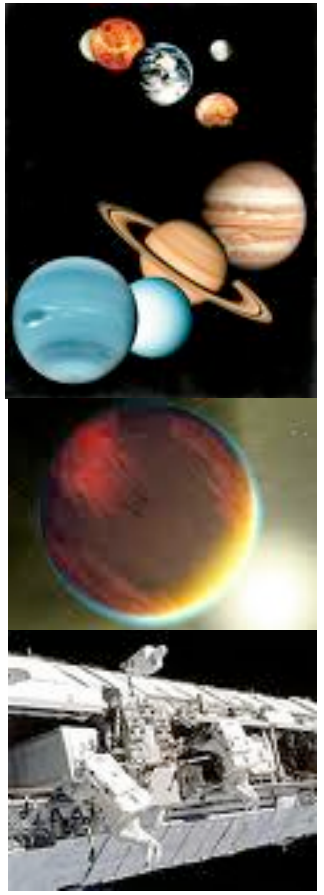
- Over-scheduling space crew time is a serious issue
- Early operational specifications - duty schedules were assumed to be long and arduous and therefore schedule were tightly packed…
- …but heavy workload was not practical for long duration missions, e.g. Skylab 4 crew rebellion due to over work
- The planned workload on Mir had to be reduced*,**
- Balance between engaging work and free time is important for crew morale
- Boredom becomes greater as mission duration lengthens
- Eight hour/day and five day/week could be considered optimal*,**



*Connors, M. M., Harison, A.A., & Atkins, F.R.(1985). *Living Aloft: Human Requirements for Extended Spaceflight, pp. 119, 123, 124, 126, 128,* 131, 172, 290, 323, NASA SP 485.
**Freeman, M. (2000). *Challenges of Human Space Exploration, pp. 8, 9, 12, 180, 188, 220, Springer,* London.
*** Levine, A. (1991). Psychological effects of long-duration space missions and stress ameliorationtechniques. In AA Harrison, YA Clearwater, & CP McKay (Eds.), From Antarctica to outer space:Life in isolation and confinement (pp. 305-315). New York: Springer-Verlag.

Space Mission Environment and Crew Work Guidelines

- Space missions workload is variable
- Crew time might not be a limited resource during transit to and from Mars but *could* be for ISS*
- ISS crew loading plan anticipates high demand for crew time
- The specified crew work time per day is:
  - 8 hours per crewmember for five days each week
  - 2 hours for exercise, 3.5 for meals, 8.5 for sleep, 1.5 personal time, 0.5 for ground coordination and planning
  - 80 minutes for payloads operations during off-duty day and days when utilization time is otherwise unavailable, and 4 hours per crewmember of station cleaning chores will be accomplished on off-duty days
- Intent of these guidelines is to prevent overloading the crew



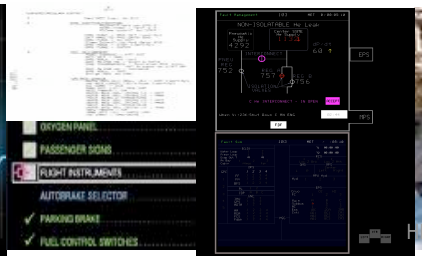
*Levri, J.A., Vaccari, D.A., & Drysdale, A.E. (2000). Theory and Application of the Equivalent System Mass Metric. SAE Paper No. 2000-01-2395, 30th International Conference on Environmental Systems, Toulouse, France, July 10-13.

**SSP 50391, Crew Loading Report, International Space Station Program, Baseline, July 1999.

Human Systems
integration division

Some questions…

- At what phases and for what questions are tools like Cooper-Harper, Bedford, or the TLX best suited to answer?
- How can interfaces, procedures, and schedules be designed from an *integrated system perspective* such that *acceptable workload* is maintained throughout the mission?
- How can workload measures be scaled up from individual task performance metrics to long duration performance?
- How are *workload requirements verified*?
- How can workload measurement devices be fielded and used appropriately for long duration mission operations?



Human Systems
integration division

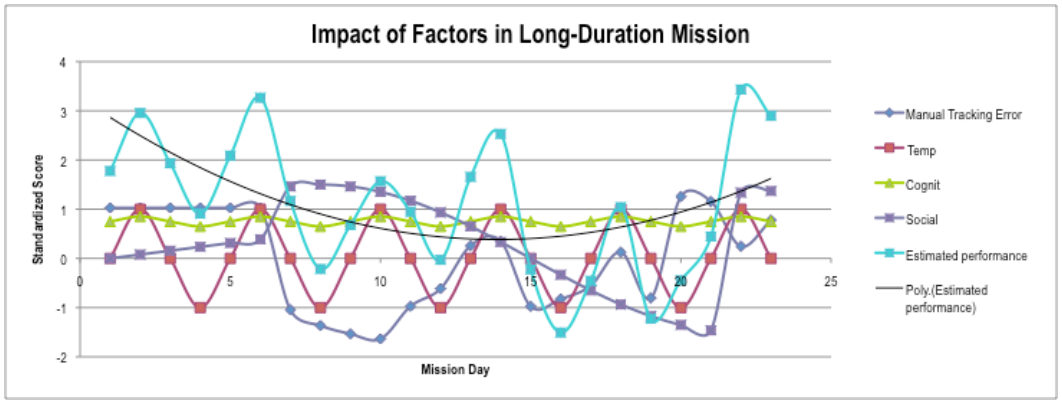
Some answers…

- Multidimensional tools (TLX) are valuable for assessing relative designs, uni-dimensional tools (with anchors) are valuable for assessing requirements
- "Optimal" workload is operator / context specific, expertise, training, time/transience of task, length of mission influences workload and its perception, that in turn can impact performance
- Requirements verification generally require a threshold measure
- Workload from the long duration mission perspective may either:
  - require a multidimensional approach OR
  - May require a new definition – i.e. we should not consider workload in the way that we have been thinking about workload

Human Systems integration division

Progress and Future Work

- Publications:
  - Gore, B.F., Macramalla, S., & Salud, E. (2010). A workshop on workload scales, measurement, and management for long duration space operations (No. NASA/CP-2010-216398). Washington, D.C.: National Aeronautics and Space Administration.
  - Casner, S.M., & Gore, B.F. (2010). Measuring and evaluating workload: A primer (No. NASA/TM 2010-216395). Washington, D.C.: National Aeronautics and Space Administration.
- Future Work:
  - Identifying the parameters that drive workload for long duration space system operations
  - Boredom will be a driver
  - Cross cultural issues
  - Gender mix…..



Human Systems
integration division

Thank you for your attention!

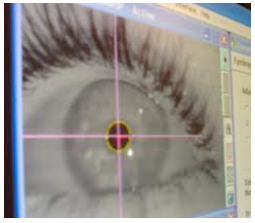
Brian F. Gore, PhD
SJSURF/NASA Ames Research Center
Human Systems Integration Division
Moffett Field CA 94035-0001
brian.f.gore@nasa.gov
650-604-2542

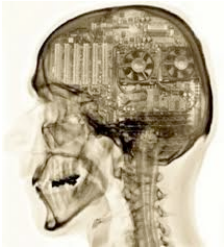
Human Systems
integration division

Extra Slides

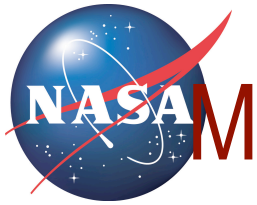
Objective Workload Measures

- Direct measures
  - Speed accuracy, Activity, Task analysis
- Indirect measures
  - Performance on a secondary task concurrently with a primary task
  - Measures the "spare capacity" of the operator
    - Operator performs a second task concurrently with the primary task, then the primary task burdens the operator with only a low or moderate amount of workload
    - Primary task performance breaks down while performing a secondary task, primary task absorbed the operator's resources, the operator is nearing the peak of his capacity to do work
  - Insensitive to the state or condition of the operator while he or she performs the task
  - Difficult to identify when operators decide to neglect the primary task in favor of the secondary task
  - Difficult to interpret - does the secondary task overlap with a given primary task; different operators can have different skill levels, or use different strategies to perform either the primary task, secondary task, or the combination of the two tasks



Casner & Gore, 2010



Motivation: Workload and Requirements Verification

- As a result a three year effort undertaken to:

1. Examine the available research on systems, on workload measurement and management, long duration operations
2. Conduct a workshop to get the most current thoughts on workload measurement and management for long duration mission operations from the system perspective
3. Develop and refine a conceptual model of parameters that impact system workload for long duration mission operation

Human Systems integration division

Subjective Measures: Bedford Scale

- Modification of the Cooper-Harper scale (aircraft handling), a 10-point scale with verbal anchors
- Operators navigate through a hierarchy, narrow down workload rating to 2 or 3 choices and then select a single rating
- Ease of use, anchors
- Can not be done concurrently with primary task
- Not an interval scale (i.e., 6 does not represent twice the workload rating of 3)
- Hierarchy is often not used, subjects proceed directly to the ten ratings
- Notion of “spare capacity” is ambiguous (is it additional time, additional mental capacity, a free hand, etc?)