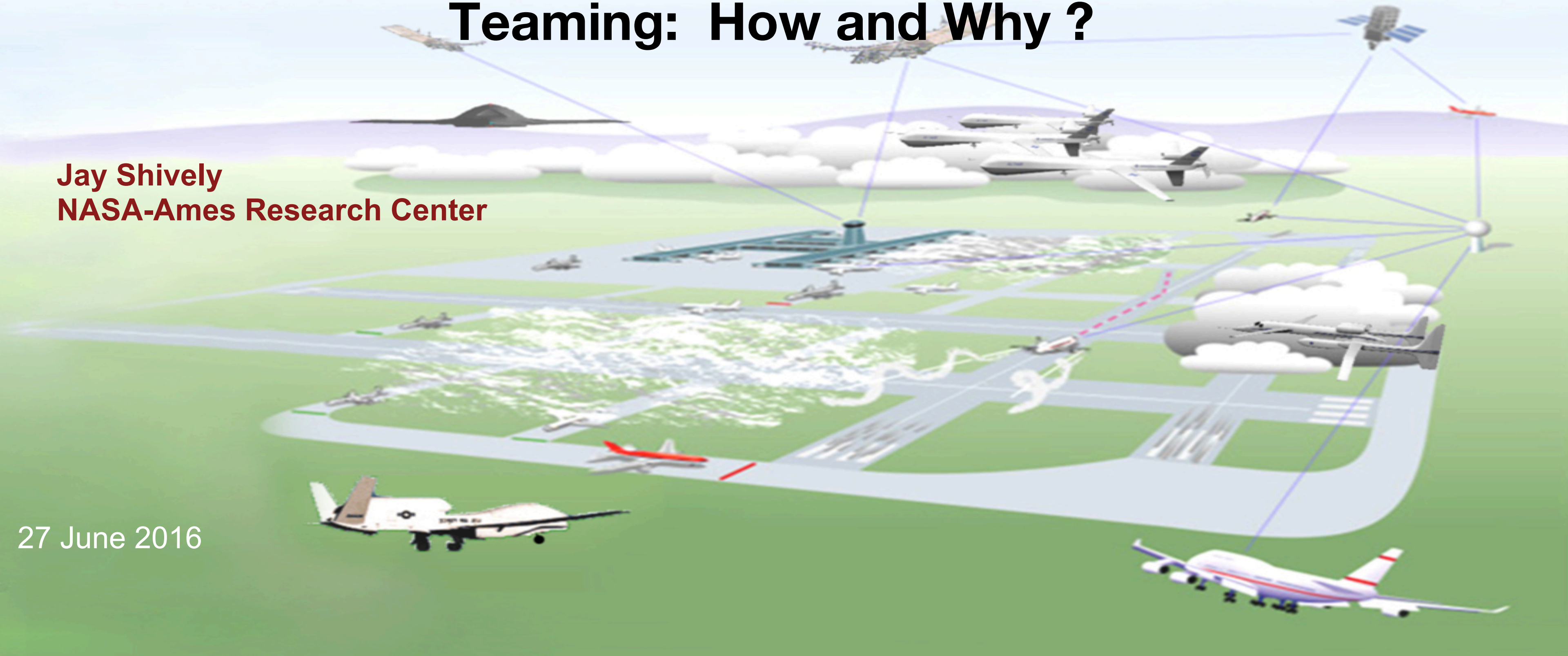
National Aeronautics and Space Administration

Workload Measurement in Human Autonomy Teaming: How and Why ?

Jay Shively
NASA-Ames Research Center

27 June 2016



Outline

- Me and my biases
  - NASA-TLX
  - MIDAS
- Categories of metrics
  - Subjective
  - Objective
  - Physiological
  - Computational
- Workload and Human Autonomy Teaming
  - Changes workload
- Right tool for the right question
  - Assessment
  - Prediction
  - Design

Jay

- Worked with Sandy Hart and the TLX team from 1984 until about 1990

- Managed the Man-Machine Integration Design and Analysis System (MIDAS) 1990 – 1995

- Bias ?

- Yes, but not in the way you might think --- I know where the skeletons are buried !!

NASA-TLX

- NASA Task Load Index (TLX)
  - Includes six subjective dimensions of global workload
  - First validation and development studies done by Hart and Staveland (1988)
  - Studies have used NASA-TLX along with physiological measures
    - Borghini et al. (2012) conducted a study assessing workload during driving a vehicle using EEG, along with other physiological data
    - Participant given NASA-TLX questionnaire at the end of each condition for subjective workload assessment
    - Correlation between NASA-TLX scores and physiological data
  - Sometimes has strongest effect size out of other workload measures
  - Sensitive to both task type and dual tasking (Matthews, Reinerman-Jones, Barber, & Abich IV, 2015)
  - Considered to be more favorable for subjects as compared to other measures of workload (Cao et al., 2009)

NASA-TLX Example

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

NASA-TLX Example

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 task or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	Good/Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing these goals?
EFFORT	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	Low/High	How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

SWAT

- Developed by Reid & Nygren (1988)
- Develops a single, global rating scale with interval properties (Rubio et al., 2004)
- Used in a variety of task settings (Hendy, Hamilton, & Landry, 1993)

I. Time Load

1. Often have spare time. Interruptions or overlap among activities occur infrequently or not at all.
2. Occasionally have spare time. Interruptions or overlap among activities occur infrequently.
3. Almost never have spare time. Interruptions or overlap among activities are very frequent, or occur all the time.

II. Mental Effort Load

1. Very little conscious mental effort or concentration required. Activity is almost automatic, requiring little or no attention.
2. Moderate conscious mental effort or concentration required. Complexity of activity is moderately high due to uncertainly, unpredictability, or unfamiliarity. Considerable attention required.
3. Extensive mental effort and concentration are necessary. Very complex activity requiring total attention.

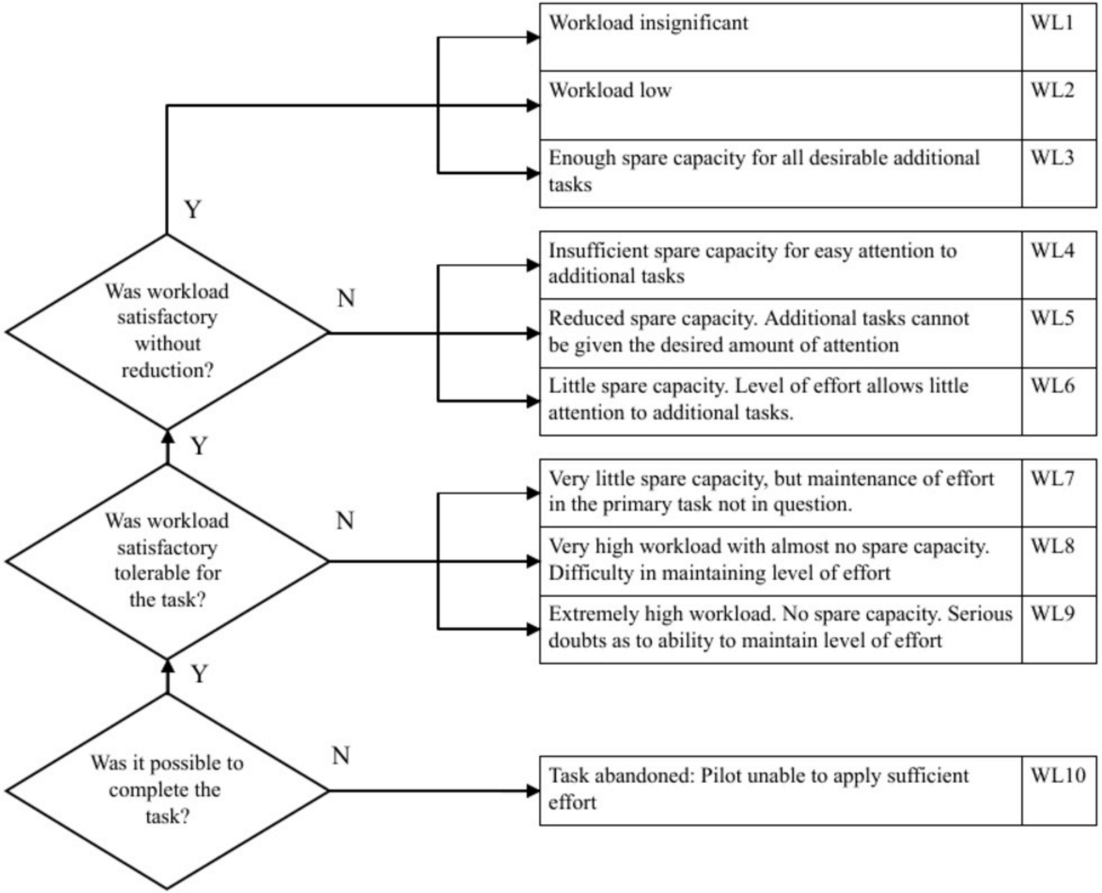
III. Psychological Stress Load

1. Little confusion, risk, frustration, or anxiety exists and can be easily accommodated.
2. Moderate stress due to confusion, frustration, or anxiety noticeably adds to workload. Significant compensation is required to maintain adequate performance.
3. High to very intense stress due to confusion, frustration, or anxiety. High extreme determination and self-control required.

Bedford Scale

- Modification of the Cooper Harper Scale (Roscoe, 1984)
- Unidimensional
- Uses a decision tree and examines whether:
  - The task can be successfully completed
  - The level of workload experienced was tolerable
  - The level of workload was satisfactory without reduction
- Taps into operator's spare mental capacity
- Currently very few studies have used it in controlled settings
- More often used in applied settings
- Not enough data on validity of the scale available (NATO Guidelines on Human Engineering Testing and Evaluation, 2001)

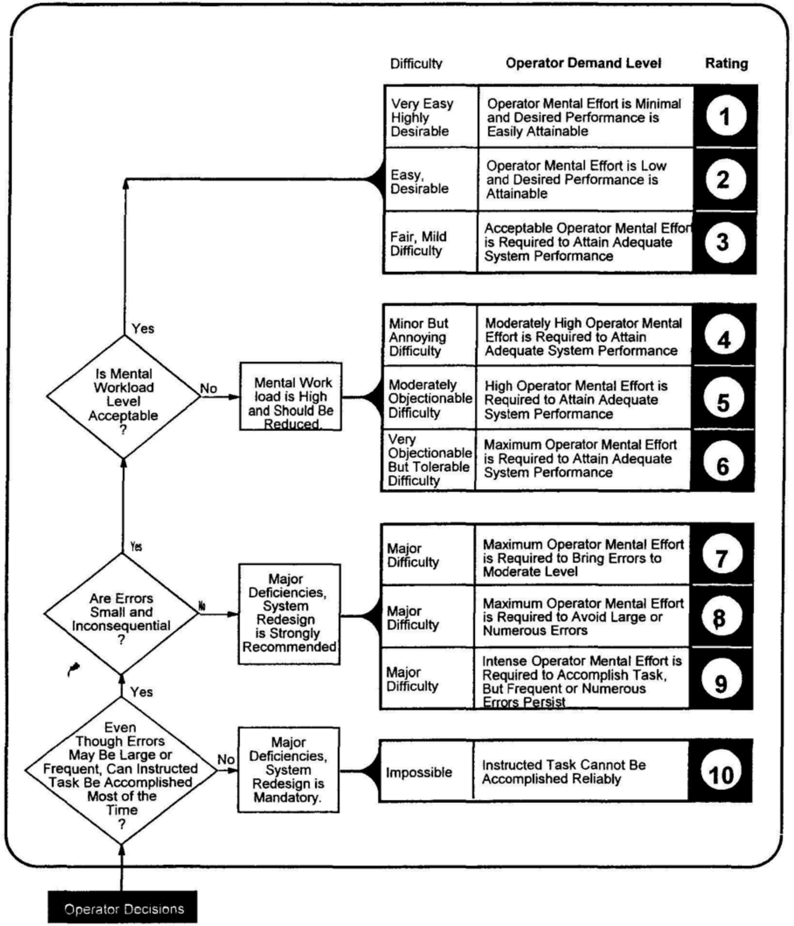
Bedford Scale



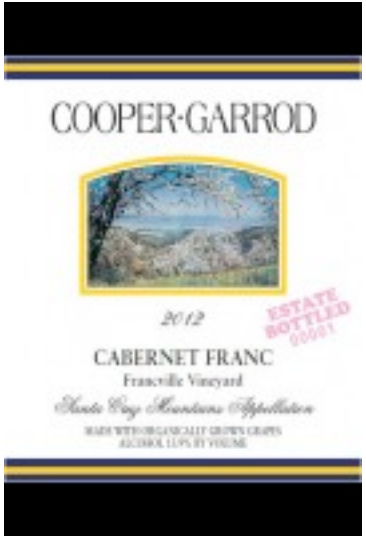
Modified Cooper Harper Scale

- Used most often in aviation
- Unidimensional
- Uses decision tree rating scale, with a score of 1 indicating "best" and a score of 10 indicating "worst"
- Relatively sensitive to changes in workload (Wierwille & Connor, 1983) and various types of workload
- Data is collected after the trial, participants poor at recalling past mental events (Wierwille & Casali, 1986)
- Limited to manual control tasks

Modified Cooper-Harper Scale



Cooper-Garrod Estate Vineyards



Subjective

Pros:

- Pretty good idea of operator's experience of workload (Crabtree, Bateman, & Acton, 1984)
- Cheap and easy (Stanton, Salmon & Walker, 2007)
- Years of use = validation (?)
- Gold standard ???
- Diagnostic when combined with objective measures (Crabtree, Bateman, & Acton, 1984)

Cons:

- Phenomenon has to be available for introspection (see SA) (Yeh & Wickens, 1984)
- Retrospective, i.e., not real-time
- Memorial (prone to memory failure ?) (Muckler & Seven, 1992)
- Not continuous, reflect average or peak
- Subjective – NOT objective data

NASA-TLX Bias

- Almost any of these (or a scale 1 – 100) gives you a pretty good idea of overall workload experienced by the operator
- NOT really diagnostic – I know of no system design ever modified because of too high "physical workload"
- Individual differences "weightings" reduce variance – but mathematically have to !!
- No one knows what the "own performance" scale means – maybe Sandy
- Have effectively become the gold standard against which other metrics – such as physio or computational models are judged

Issues of Subjective Measures

- **Processing characteristics are lost** (Yeh & Wickens, 1984)
  - Limited in scope
  - Only provides scalar measures
- **Dissociations between subjective and objective measures** (Yeh & Wickens, 1988)
- **Difficult to compare results across scales** (Gopher & Braune, 1984)
  - Lack of formal theory for workload
  - Subjective measurement scales are influenced by how experimenters select scalar dimensions for rating

Objective

Metrics:

- Embedded Secondary tasks
- Naturally occurring secondary tasks
- WL probes

Workload Probes

- **Situation Present Assessment Method (SPAM)**
  - On-line probe method that can measure workload, in addition to SA (Stanton, Salmon, & Walker, 2007)
  - **Readiness latency**: Time from onset of “ready” prompt for query to an individual’s response to the prompt acts as an indicator for workload
    - Objective (Stanton, Salmon, & Walker, 2007)
    - Sometimes accompanied by an auditory warning signal (Pierce, 2012)
  - Queries can also ask operator to report current mental workload (Silva et al., 2013)
    - Scale
    - Subjective rating
  - Not intrusive to operator performance and workload (Silva et al., 2013; Keeler et al., 2015)

Embedded Secondary Tasks

- Operator performs a primary task in addition to a secondary task
- A wide variety of secondary tasks have been used in studies (Ogden, Levine, & Eisner, 1979)
- RT on secondary tasks often shows greatest sensitivity to workload changes
- Time estimation task is also sensitive, but can be intrusive (Wierwille, Rahimi, & Casali, 1985)
- Change detection also successful and less intrusive (Teo, Reinerman-Jones, & Szalma, 2015)

Objective

Pros:

- Objective data: RT, error

Cons:

- Spacific to each implementation
- Low data rate
- Can be difficult to implement
- Momentary measure (not continuous)
- Can’t implement in some situations (real cockpit- can’t add secondary tasks)

Physiological

Metrics:

- Heart-rate (variability)
- Eye gaze
- GSR
- Eye-blink
- fMRI
- EEG
- fNIRS

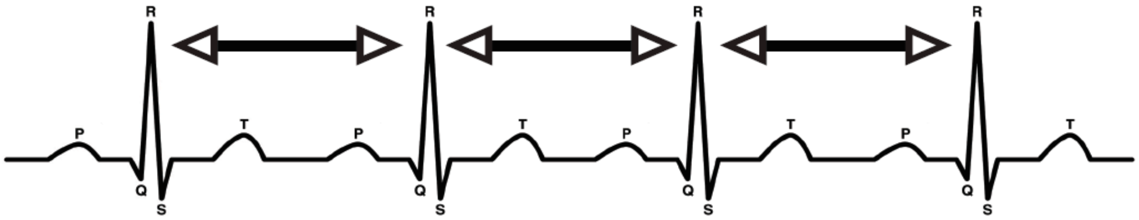
Physiological Measures

- **Eye Gaze**
  - Gold, Körber, Lechner, and Bengler (2016)
    - Study to determine how traffic density and verbal tasks affect take-over performance in highly automated driving
      - I.e. how much time does the driver need in order to regain control of an automated vehicle if a situation necessitates this? Does traffic density and a verbal task have an affect on performance of taking back control?
    - Used eye gaze behavior as a measure of workload
      - Lower horizontal gaze distribution (HGD) = More workload (Wang, Reimer, Dobres, & Mehler, 2014)

Physiological Measures

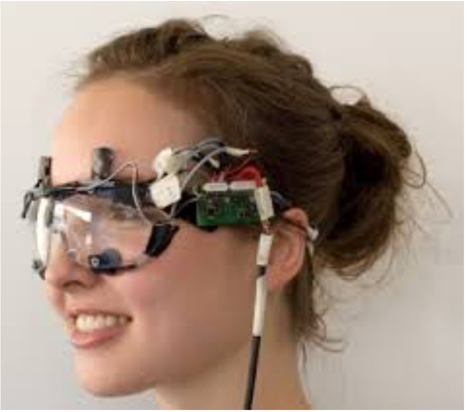
- **Heart Rate Variability**
  - Decrease in heart rate variability may indicate an increase in mental workload (Mulder, 1980)
  - Strang, Best, and Funke (2014)
    - Studied mental workload of participants in a simulated training exercise involving realistic, large-scale air-combat scenarios.
    - Examined the ability of heart rate to predict mental workload.
    - Some data to support that heart rate may be able to predict mental workload, but this relationship is inconsistent.

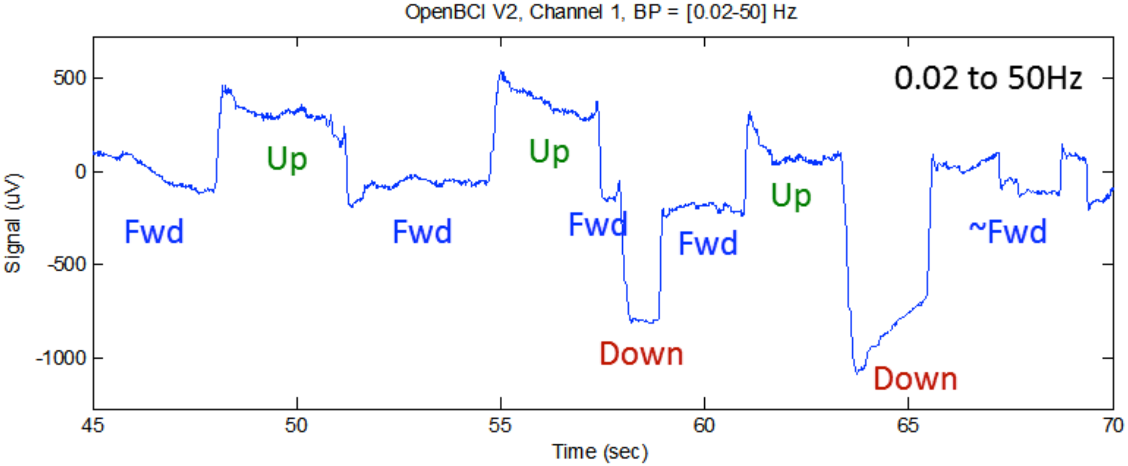
Figure 1: Heart rate variability is measured by calculating the time between R spikes on an ECG trace



Physiological Measures

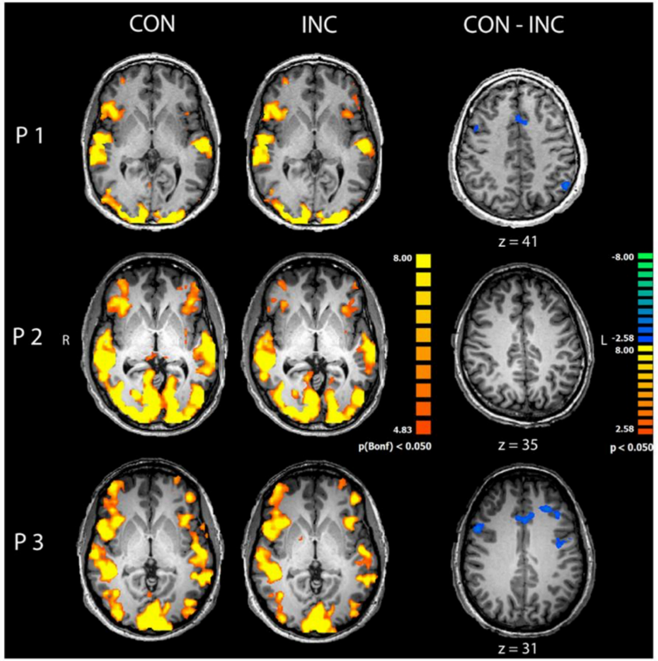
- **Eye Blink**
  - Sometimes measured through electrooculography (EOG) (Veltman & Gaillard, 1996)
  - Length or frequency of blink
  - Not always sensitive to changes in workload (Wierwille & Connor, 1983)
  - Might need to be combined with other eye tracking techniques to be more reliable (Orden, Limbert, & Makeig, 2001)





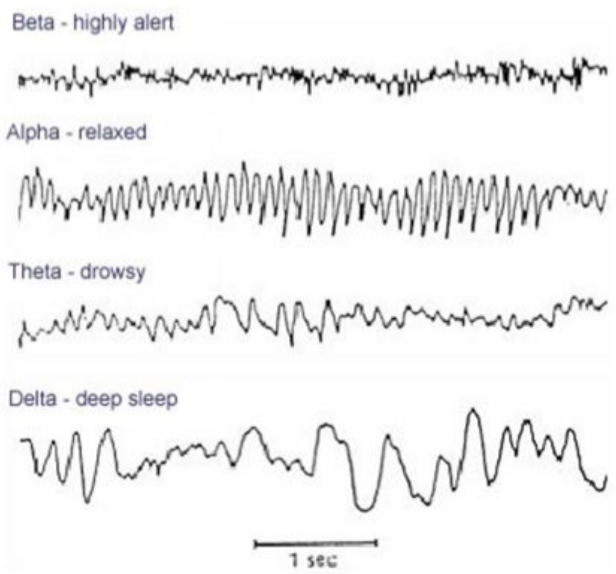
Physiological Measures

- **Functional Magnetic Resonance Imaging (fMRI)**
  - Monitoring cerebral blood flow velocity (CBFV)
  - As CBFV increases in the prefrontal cortex, mental workload increases (Parasuraman & Caggiano, 2005)
  - Highly constrained environment
  - Limits what kind of activities can be analyzed (Warm, Parasuraman, & Matthews, 2008)



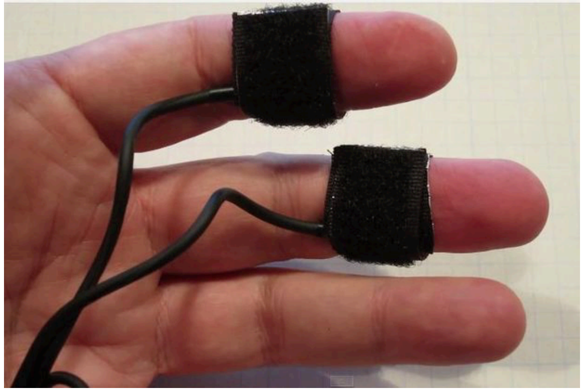
Physiological Measures

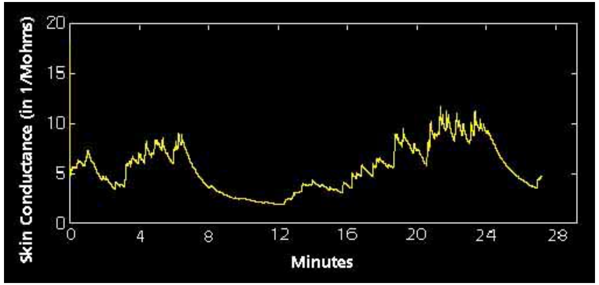
- **Electroencephalogram (EEG)**
  - **Electrodes are placed on the scalp over various brain areas:**
    - Fz, F3, F4, Cz, C3, C4, Pz, P3, P4
  - Different types of brain waves
    - Alpha (7-14 Hz)
    - Beta (14-30 Hz)
    - Theta (4 to 7 Hz)
    - Delta (up to 4 Hz)
  - As mental workload increases, alpha waves are replaced by beta waves, and frontal theta waves are increased (Borghini, et al., 2012)



Physiological Measures

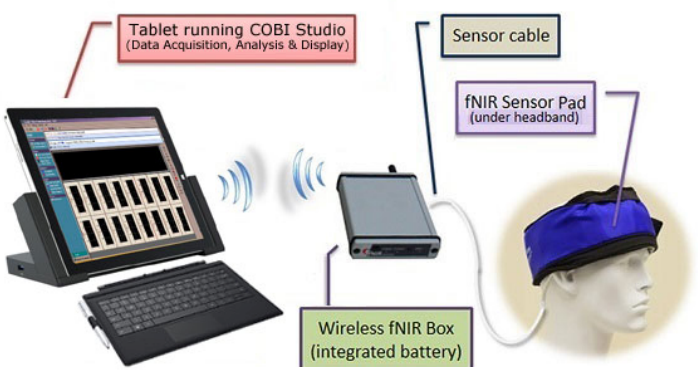
- **Galvanic Skin Response (GSR)**
  - Measurement of resistance of skin tissue to electrical current
  - Measured through palms, inside / outside of wrist, arch of foot, forehead, or fingers
  - Susceptible to individual differences in response (Wierwille, 1979)
  - Found to be associated with cognitive workload (Shi et al., 2007)
    - Mean GSR increases as cognitive load increases

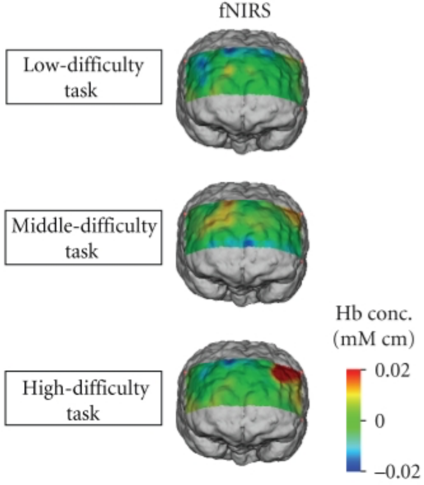




Physiological Measures

- **Functional Near-Infrared Spectroscopy (fNIRS)**
  - Relatively new measure
  - Monitors elevation of rSO_2
  - Higher rSO_2 levels = more cognitively demanding (Matthews, Renierman-Jones, Barber, & Abich IV, 2015)
  - Not always sensitive enough to changes, but does correlate well with other physiological measures of workload (e.g. HR) (Teo, Reinerman-Jones, & Szalma, 2015)





Physio

Pros:

- Continuous
- Potentially unobtrusive
- Face Validity – looks scientific (Levin et al., 2006)
- Supplements subjective measures (Wierwille & Eggemeier, 1993)

Cons:

- Not a "pure" workload signal
- Individual differences (Wierwille, 1979)
- Sensitive to external events/sources
- Poor correlation to subjective metrics (gold standard)

Computational

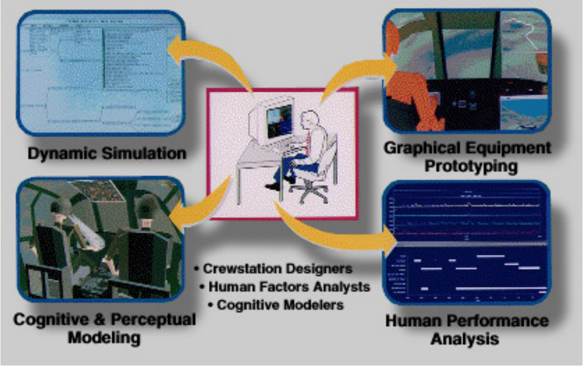
Models

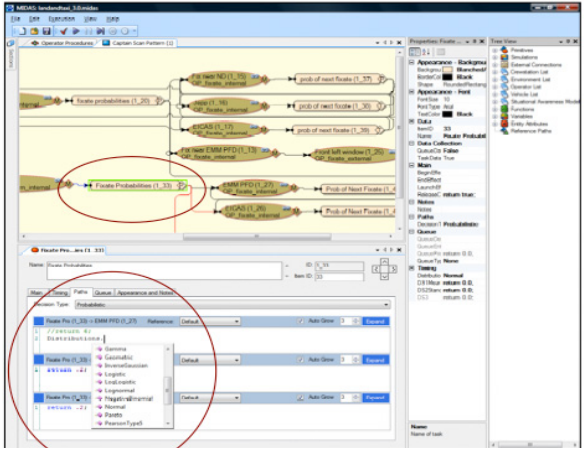
- MIDAS
- IMPRINT
- OMAR
- ACT-r*

Human Performance Modeling in Aviation, Foyle, D.C. & Hooey, B.L. (2008)

MIDAS

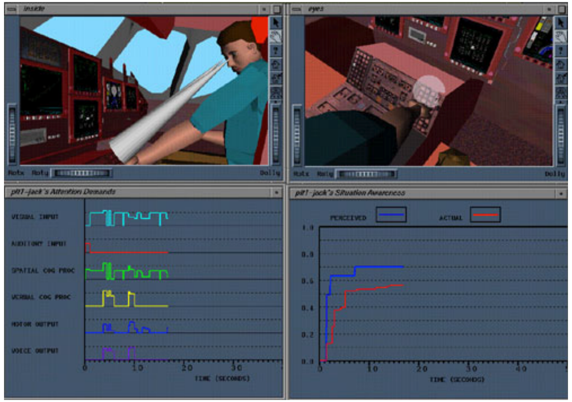
- Man-machine Integration Design and Analysis System (MIDAS)
- NASA Ames Research Center human performance model (HPM) software tool
- Predict human-system performance and model human error
- “What-if” analyses (Gore, 2011)
- 3-D rapid prototyping, dynamic simulation, and human performance modeling with the aim to reduce design cycle time
- Links a virtual human to a computational cognitive structure that represents human capabilities and limitations (NASA, 2016)
- Currently v5 being used

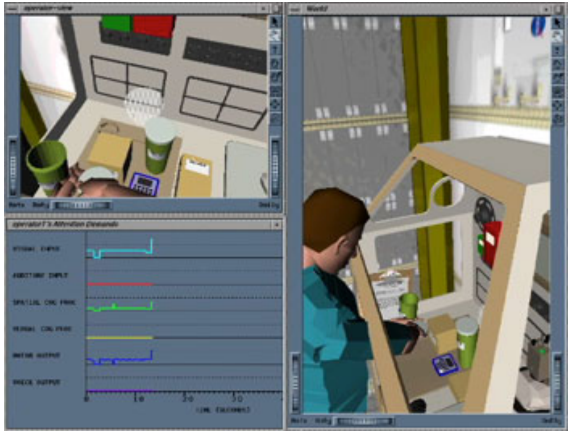




MIDAS

- Cognitive component
  - Perceptual mechanism
  - Memory
  - Decision Maker
  - Response selection architecture
- Outputs include:
  - Dynamic visual representations
  - Timelines
  - Task lists
  - Cognitive loads along resource channels
  - Actual / perceived S.A.
  - Human error vulnerability
  - Human performance quality





MIDAS Architecture

Inputs

Crew Station Design, Cockpit Geometry, Display Control Layout, Cockpit Equipment Functionality

Mission / Tasks, Flight Profiles, Waypoints, Other Scenario Objects, Planned Operator Mission Activities

Operator Characteristics, Cognitive Physical, Motor Response Parameters

User Interface

Interactive Graphical Interface

Cockpit Design Editor

Equipment Editor

Route Editor

Activity Editor

Interactive Cockpit Design Tools

Jack
Anthropometric Model

Vision Models

Simulation System Models & Tools

Symbolic Operator Model
Vision, Perception, Attention, World Representation, Decision, Scheduling, Task Loading, Mission Activities, Motor

World Model
Vehicle, Cockpit Equipment, Flight Dynamics

User Interface

Interactive Graphical Interface

Display Animation

Runtime Data Graphical Displays

Summary Data Graphical Displays

3D Graphical Display

Outputs

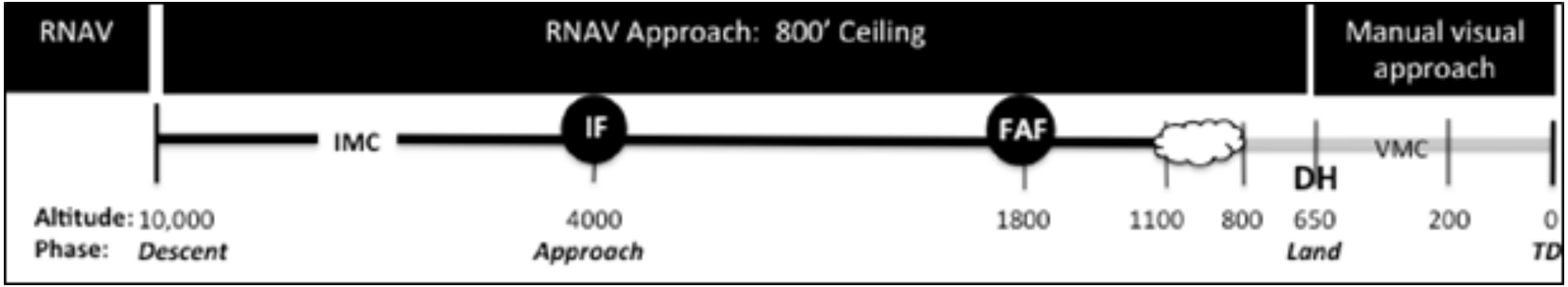
Ergonomic Analysis Results
Reach, Fit, & Other MIL-STD Analysis, Visibility and Legibility

Mission Operator Performance Measures
Accuracy, Info FLow, Response Times, Activity Traces, Task Load Timelines, Resource Conflicts

Visualization of Sim. Missions Operator Activities, Equipment Status

MIDAS

- Air MIDAS
  - Aviation specific version of MIDAS
  - Models ATCo
  - Switches between control strategies depending on number of aircraft under control and the complexity of maneuvers the aircraft have to perform



IMPRINT

- Improved Performance Research Integration Tool (IMPRINT)
- Developed by the U.S. Army Research Lab, Human Research & Engineering Directorate
- Software is available for free for
  - U.S. government agencies
  - U.S. private industry with U.S. government contract
  - U.S. colleges and universities working in HSI
- Designed to support system performance through lifecycle
- Can help to
  - Set realistic system requirements
  - Identify Soldier-driven constraints on system design
  - Evaluate the capability of available manpower and personnel to effectively operate and maintain a system under environmental stressors (U.S. Army, 2016)

IMPRINT Modules

Warfighter

Estimate the type of individuals who will be available to operate and maintain the system

Missions

Estimate the effect of operator performance on system performance, including time, accuracy, or mental workload

Equipment

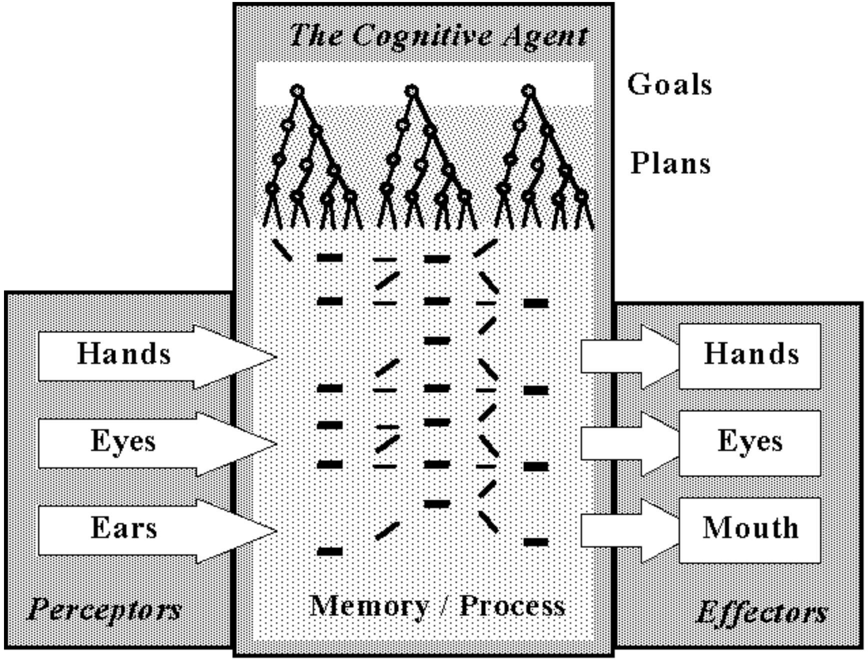
Estimate maintenance man-hours required to attain acceptable system availability

Forces

Estimate the manpower needed to complete the routine and unplanned work performed by a force unit

Operator Model Architecture (OMAR)

- Provides a simulation environment that allows for modeling human operators, where they work, and the entities of the larger world that are reflected in their workplaces
- A "production rule-based executive process" regulates scheduling of competing tasks
- Emphasis on developing multiple-task behaviors from "functional centers" that are operating at the same time without an executive or central control (Deutsch, 1998)



Unique Characteristics of OMAR

- Stimuli directly affect procedural memory
- "Function-specific procedures" that represent specific brain areas coordinate the completion of tasks
- Resulting behaviors may be considered "intelligent"
- Task completion is mediated on a pair wise basis and not through a central executive. (Deutsch, 1998)

Computational

Pros:

- Learn a lot by formalizing description of your system
- Objective (sort of – input, assumptions, etc.)
- What if questions can be asked
- Can mod and re-run
- Consistent
- Can be done with notional system

Cons:

- Benefit might largely be in the process (more of my bias)
- Need a detailed task analysis/system design
- !@$#% input > !@#$ output

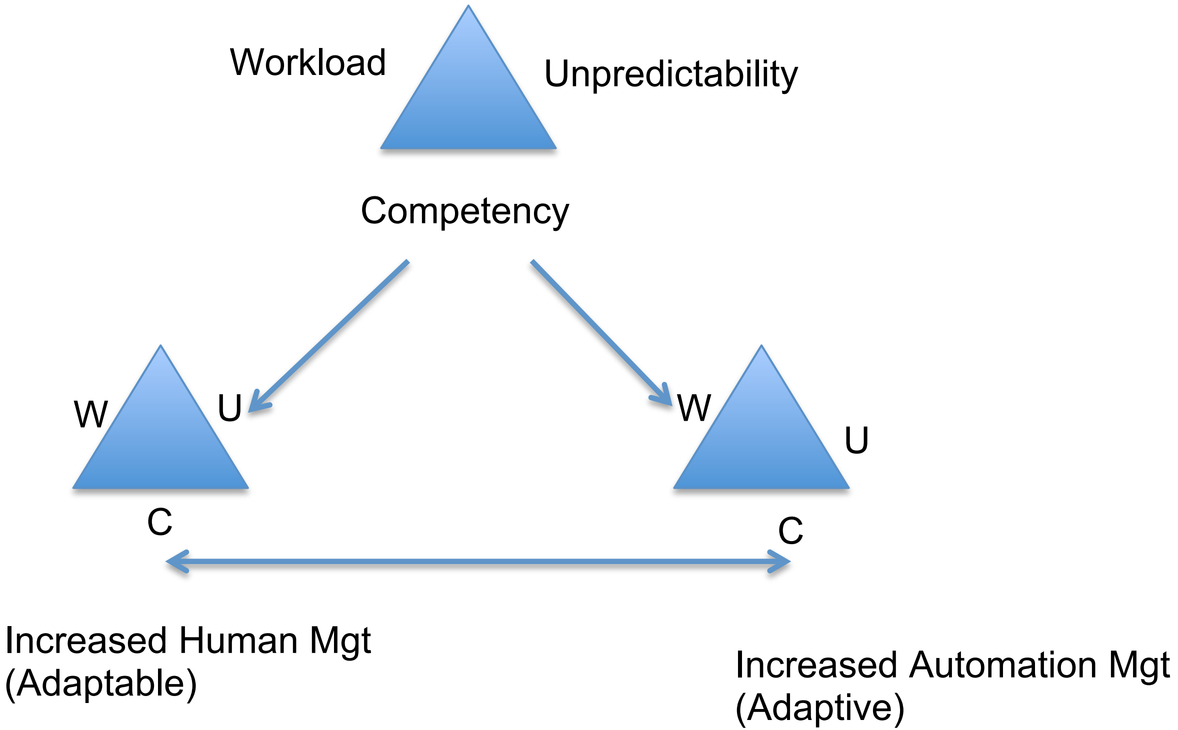
Relationship Between Automation & Workload

- Automation does not necessarily reduce workload, **just changes it.**

- Automation changes an operator's role from manually **controlling** a system to **monitoring** the automated system (Parasuraman & Riley, 1997)

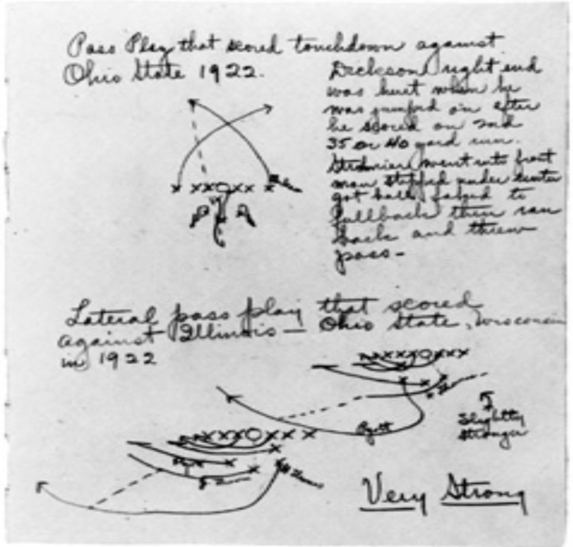
- **Examples**
  - **Wiener (1989)**
    - Pilot responses were divided when asked whether workload was decreased in a more automated cockpit
  - **Warm, Dember, & Hancock (1996)**
    - Monitoring tasks can lead to underarousal and increased mental workload
  - **Wiener & Curry (1980)**
    - Although automation may reduce *manual* workload, it may increase *overall* workload as a result of increased mental workload.

Relationship Between WL and Automation



Miller, C.A. & Parasuraman, R. (2007)

A Playbook® Approach to Delegation



A page from Alonzo Stagg's 1927 Playbook

- A means of Delegation
- Plays contain an implicit goal
- Plays define a "template" of plan/behavior alternatives—a "space" of delegated planning authority
  - "pre-compiled" with convenient label
  - Supervisor can further constrain/stipulate as desired– by reference to play structure
  - Monitoring and information reporting facilitated by shared intent structure
  - Dynamic, real time revision and tuning = "calling signals"
- Subordinates responsible for best-effort attempts within play constraints

Playbook & HAT

- Single Operator control of multiple, heterogeneous UAS

(Simulations and flight tests)

  - Top ten pre-defined Plays – from operators
    - Convey support
    - Troops in contact
    - Recon an area
  - Increased Performance
  - Decreased Workload
- Human Autonomy Teaming
  - Reduced Crew Compliment in Commercial Aviation
  - One step further – not just delegation, but discussion, negotiation, joint problem solving
  - Automation (and interface) adapts by (largely) pilot-directed context

Why measure workload ?

System Life-cycle

- Design
- Evaluation (R & D)
- Evaluation (Operational)
- Embedded (adaptive automation)

- WC Fielde: Workload Consultant for Field Evaluation

Design

Environment:

- System doesn't exist
- SME's may be tangential
- Non-real time

Decisions:

- Roles and responsibilities
- Information flow/ displays
- Crew size

Metric:

Computational Models

Evaluation (R&D)

Environment:

- Prototype system
- Focus on other variables
- Real-time

Questions:

- Workload too high/low
- Effect of variables on WL

Metrics:*

- Subjective
- Objective/secondary
- Phsyio

* Choice depends on ability to insert/identify secondary tasks

Evaluation (Operational)

Environment:

- System
- Real users
- Real-time

Questions:

- Workload too high/low

Metrics:

- Subjective
- Physio (if non-intrusive)

Embedded (e.g., Adaptive Automation)

Environment:

- System (WL eval is part of the system)
- Real users
- Real-time

Questions:

- Workload too high/low

Metrics:

- Subjective
- Performance
- Physio (if non-intrusive)

Summary

- Pros and Cons of all approaches
- Driven by the QUESTION
- Strongly advise using a battery of measures to converge on "workload"

- Adaptable vs. Adaptive Automation…

References

Borghini, G., Vecchiato, G., Toppi, J., Astolfi, L., Maglione, A., Isabella, R., Caltagirone, C., Kong, W., Wei, D., Zhou, Z., Polidori, L., Vitiello, S., Babiloni, F. (2012). Assessment of mental fatigue during car driving by using high resolution EEG activity and neurophysiologic indices. *34th Annual International Conference of the IEEE EMBS.*

Budiu, R. (n.d.). About. Retrieved June 07, 2016, from http://act-r.psy.cmu.edu/about/.

Cao, A., Chintamani, K. K., Pandya, A. K., & Ellis, R. D. (2009). NASA TLX: Software for assessing subjective mental workload. *Behavior research methods*, 41(1), 113-117.

Crabtree, M. S., Bateman, R. P., & Acton, W. H. (1984, October). Benefits of using objective and subjective workload measures. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 28, No. 11, pp. 950-953). Sage Publications.

Deutsch, S. (1998). Interdisciplinary Foundations for Multiple-task Human Performance Modeling in OMAR. *Twentieth Annual Meeting of the Cognitive Science Society.*

Fallahi, M., Motamedzade, M., Heidarimoghadam, R., Soltanian, A. R., & Miyake, S. (2016). Effects of mental workload on physiological and subjective responses during traffic density monitoring: a field study. *Applied ergonomics,* 52, 95-103.

Geddie, J. C., Boer, L. C., Edwards, R. J., Enderwick, T. P., & Graff, N. (2001). NATO Guidelines on Human Engineering Testing and Evaluation (No. RTO-TR-021). *NATO Research and Technology Organization* (France).

Gold, C., Körber, M., Lechner, D., Bengler, K. (2016). Taking Over Control From Highly Automated Vehicles in Complex Traffic Situations: The Role of Traffic Density. *Human Factors, 58*(4), 642-52. doi: 10.1177/0018720816634226.

Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures?. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 26(5), 519-532.

Gore, B. F. (2011). Man–machine integration design and analysis system (MIDAS) v5: Augmentations, motivations, and directions for aeronautics applications. In *Human modelling in assisted transportation* (p. 43-54). Springer Milan.

References

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, 52, 139-183.

Hendy, K. C., Hamilton, K. M., & Landry, L. N. (1993). Measuring subjective workload: when is one scale better than many?. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 35(4), 579-601.

Keeler, J., Battiste, H., Hallett, E. C., Roberts, Z., Winter, A., Sanchez, K., Strybel, T.Z., & Vu, K. P. L. (2015). May I Interrupt? The effect of SPAM Probe Questions on Air Traffic Controller Performance. Procedia Manufacturing, 3, 2998-3004.

Lebiere, c., Anderson, J. R., Bothell, D. (2001). Multi-Tasking and Cognitive Workload in an ACT-R Model of a Simplified Air Traffic Control Task. *Proceedings of the Tenth Conference on Computer Generated Forces and Behavior Representation, 2001.*

Levin, S., France, D. J., Hemphill, R., Jones, I., Chen, K. Y., Rickard, D., ... & Aronsky, D. (2006). Tracking workload in the emergency department. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(3), 526-539.

Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich, J. (2015). The psychometrics of mental workload multiple measures are sensitive but divergent. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(1), 125-143.

Moray, N. (1982). Subjective mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 24(1), 25-40.

Muckler, F. A., & Seven, S. A. (1992). Selecting performance measures:" Objective" versus" subjective" measurement. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 34(4), 441-455.

NASA (2016). Man-machine Integration Design and Analysis System (MIDAS). Retrieved on June 7, 2016 from http://human-factors.arc.nasa.gov/groups/midas/.

Ogden, G. D., Levine, J. M., & Eisner, E. J. (1979). Measurement of workload by secondary tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 21(5), 529-548.

References

Orden, K. F., Limbert, W., Makeig, S., & Jung, T. P. (2001). Eye activity correlates of workload during a visuospatial memory task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 43(1), 111-121.

Parasuraman, R., Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors, 39*(2), 230-253.

Parasuraman, R., & Caggiano, D. (2005). Neural and genetic assays of human mental workload. *Quantifying human information processing*, 123-149.

Pierce, R. S. (2012). The effect of SPAM administration during a dynamic simulation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 54(5), 838-848.

Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. *Advances in psychology*, 52, 185-218.

Roscoe, A. H. (1984). Assessing pilot workload in flight. Royal Aircraft Establishment. Bedford (England).

Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology*, 53(1), 61-86.

Shi, Y., Ruiz, N., Taib, R., Choi, E., & Chen, F. (2007, April). Galvanic skin response (GSR) as an index of cognitive load. In *CHI'07 extended abstracts on Human factors in computing systems* (pp. 2651-2656). ACM.

Silva, H. I., Ziccardi, J., Grigoleit, T., Battiste, V., Strybel, T. Z., & Vu, K. P. L. (2013). Are the intrusive effects of SPAM probes present when operators differ by skill level and training?. In *Human interface and the management of information. Information and interaction design* (p. 269-275). Springer Berlin Heidelberg.

Stanton, N., Salmon, P. M., & Rafferty, L. A. (2013). Human factors methods: a practical guide for engineering and design. Ashgate Publishing, Ltd.

Strang, A. J., Best, C., Funke, G. J. (2014). Heart Rate Correlates of MEntal Workload in a Large-Scale Air-Combat Simulation Training Exercise. *Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting.*

Teo, G., Reinerman-Jones, L., Matthews, G., & Szalma, J. (2015). Comparison of Measures Used to Assess the Workload of Monitoring an Unmanned System in a Simulation Mission. Procedia Manufacturing, 3, 1006-1013.

U.S. Army (2016). Improved Performance Research Integration Tool. Retrieved on June 7, 2016 from http://www.arl.army.mil/www/default.cfm?page=445.

Veltman, J. A., & Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task. *Biological psychology*, 42(3), 323-342.

Wang, Y., Reimer, B., Dobres, J., Mehler, B. (2014). The sensitivity of different methodologies for characterizing drivers' gaze concentration under increased cognitive demand. *Transportation Research Part F 26,* 227-237. doi: 10.1016/j.trf.2014.08.003

Warm, J. S., Dember W. N., Hancock, P. A. (1996). Vigilance and workload in autmated systems. *Automation and human performance: Theory and applications,* 183-200.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 50(3), 433-441.

Wiener, E. L. (1989). Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft. *NASA Contractor Report 177528.*

Wiener, E. L., Curry, R. E. (1980). Flight-deck automation: promises and problems. *Ergonomics 23*(10), 995-1011.

Wierwille, W. W., & Connor, S. A. (1983). Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 25(1), 1-16.

References

Wierwille, W. W., Rahimi, M., & Casali, J. G. (1985). Evaluation of 16 measures of mental workload using a simulated flight task emphasizing mediational activity. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 27(5), 489-502.

Wierwille, W. W., & Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 35(2), 263-281.

Wierwille, W. W. (1979). Physiological measures of aircrew mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 21(5), 575-593.

Yeh, Y. H., & Wickens, C. D. (1984). The dissociation of subjective measures of mental workload and performance.

Yeh, Y. Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 30(1), 111-120.