Aligned with your needs.

Using Operator Workload Data to Inform Human Reliability Analyses

Lila Laux and Christopher Plott

August 28, 2007

ALION SCIENCE AND TECHNOLOGY

Context – Burden of Additional Operator Actions

- There have been changes in operator workload as a result of additional demands on operations personnel to perform actions inside and outside the control room
    - Substitution of manual actions for automatic actions
    - Manual actions used in lieu of fire barriers
    - Increased communications between control room operators and grid load dispatchers to maintain grid reliability
    - Reduced response times due to power uprates
    - Reduced plant staffing
    - Workarounds
- The cumulative effect of operator workloads, such as manual actions and workarounds, on the performance of plant operations personnel and plant safety had not been examined

Background

- The goal of this research was to investigate approaches for reviewing the impact of the cumulative effects of workload on operator performance and plant safety

- Two of the key tasks were to:
  - Assess workload strategies and assessment methodologies that could support the objectives
  - Develop appropriate methods to assess the impact of operations personnel workload on plant safety

- Since this work was sponsored by the Nuclear Regulatory Commission (NRC), the methods need to:
  - Support risk-informed decision-making
  - Be suitable to license reviewers and plant inspectors

- This presentation discusses the results our initial efforts

KEY TASK: Assess workload strategies and assessment methodologies that could support the objectives

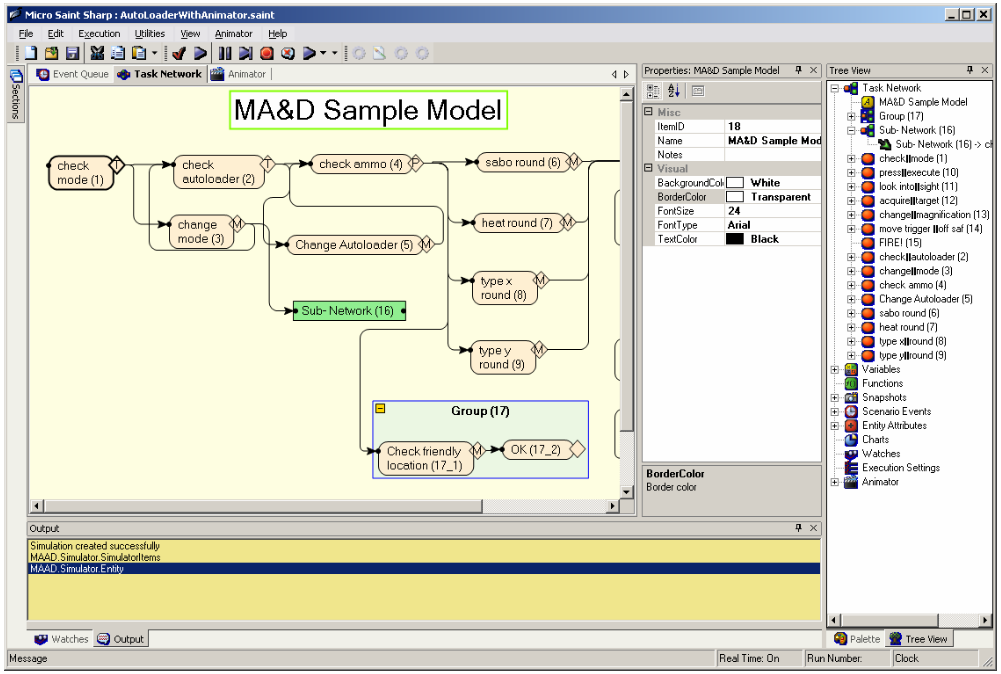
- Two key types of workload assessment techniques identified as most applicable to this effort:
  - Projective – Used to predict what workload an operator will experience. Typically implemented using human performance modeling tools.
  - Assessment – Used to subjectively self-report perceptions of workload while or after performing a task

- Methods evaluated based on:
  - Breadth of usage
  - Conceptual and application complexity
  - Manual or automated application
  - Diagnosticity (identify source of workload)
  - Validity

- VACP and NASA TLX were the respective projective and assessment methods selected for use in this research

Workload Prediction Techniques

- Typically used in Task-Network Human Performance Models
- For processes that can be described by a series of discrete activities or tasks
- For each task, connections to other tasks, branching, logics, task time and timing characteristics, and resource utilization – including workload are assigned
- When models are executed, changes in workload demand are tracked, and may be used to impact the flow of the model
- Predictive techniques
  - VACP: Visual, Auditory, Cognitive, and Psychomotor
  - W/INDEX
  - POP: Prediction of Operator Performance
  - IP/PCT: Information Processing/ Perceptual Control Theory
  - POPIP: Prediction of Operator Performance with Information Processing



VACP – Visual, Auditory, Cognitive, Psychomotor

- Four task demand channels: Visual, Auditory, Cognitive, Psychomotor
- Based on task resource demand concept of McCraken and Aldrich, 1984
- Uses interval ratings derived from the nature of the task
- Analyst or SME assigns a number to each channel for each task in structured manner
- Uses simple calculus based on sums for each channel and across channels from all active tasks
- Can also be used for assessment

VACP: Scale Descriptors

Visual

0.0 No visual activity
1.0 Visually register, detect occurrence
3.7 Visually discriminate
4.0 Visually inspect / check
5.0 Visually locate / align
5.4 Visually track / follow
5.9 Visually read (symbol)
7.0 Visually scan / search / monitor

Auditory

0.0 No auditory activity
1.0 Detect / register sound
2.0 Orient to sound (general)
4.2 Orient to sound (selective)
4.3 Verify auditory feedback
4.9 Interpret semantic content (speech)
6.6 Discriminate sound characteristics
7.0 Interpret sound patterns

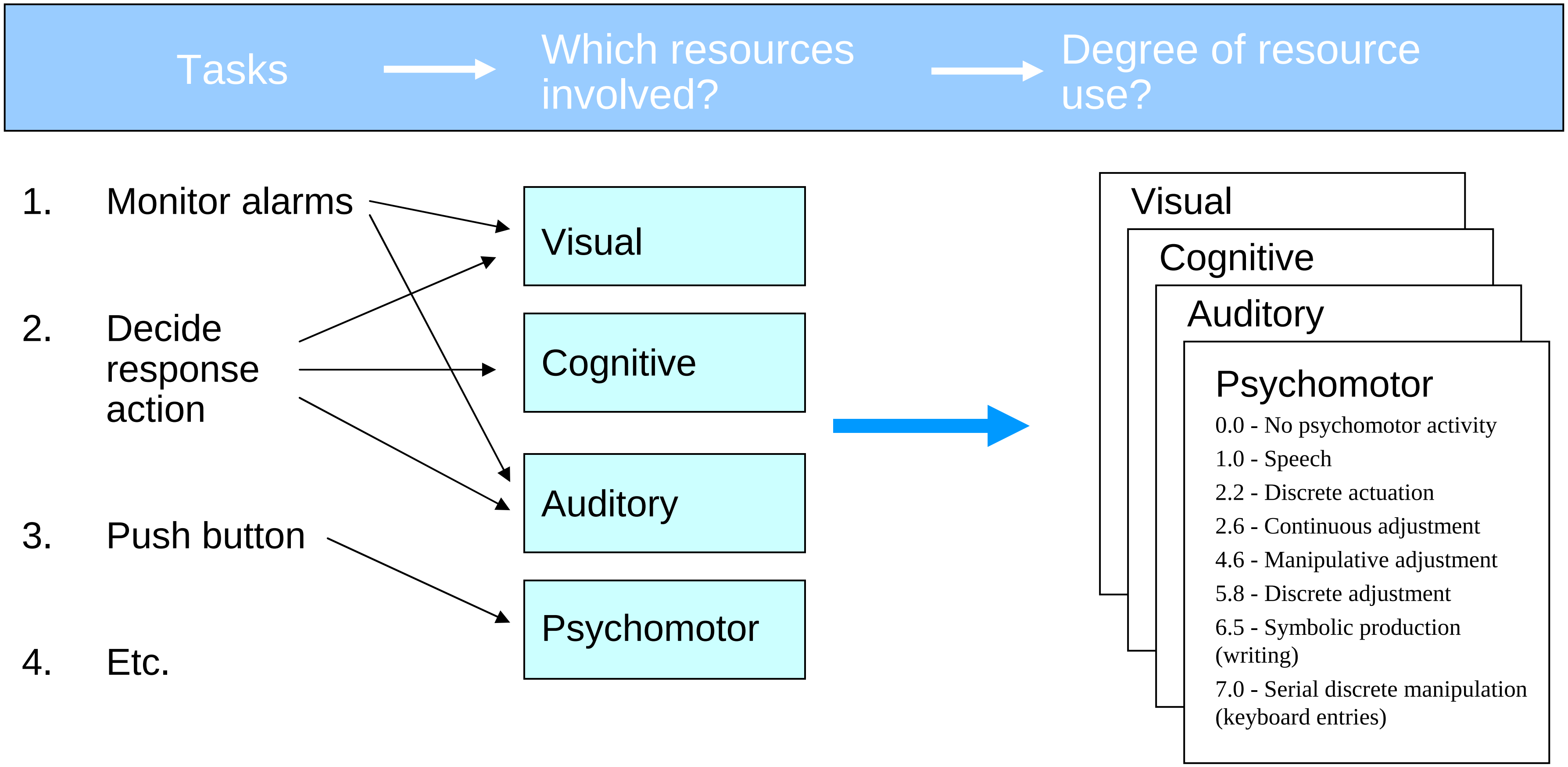
Cognitive

0.0 No cognitive activity
1.0 Automatic, simple association
1.2 Alternative selection
3.7 Sign / signal recognition
4.6 Evaluation / judgement
5.3 Encoding / decoding, recall
6.8 Evaluation / judgement
7.0 Estimation, calculation, conversion

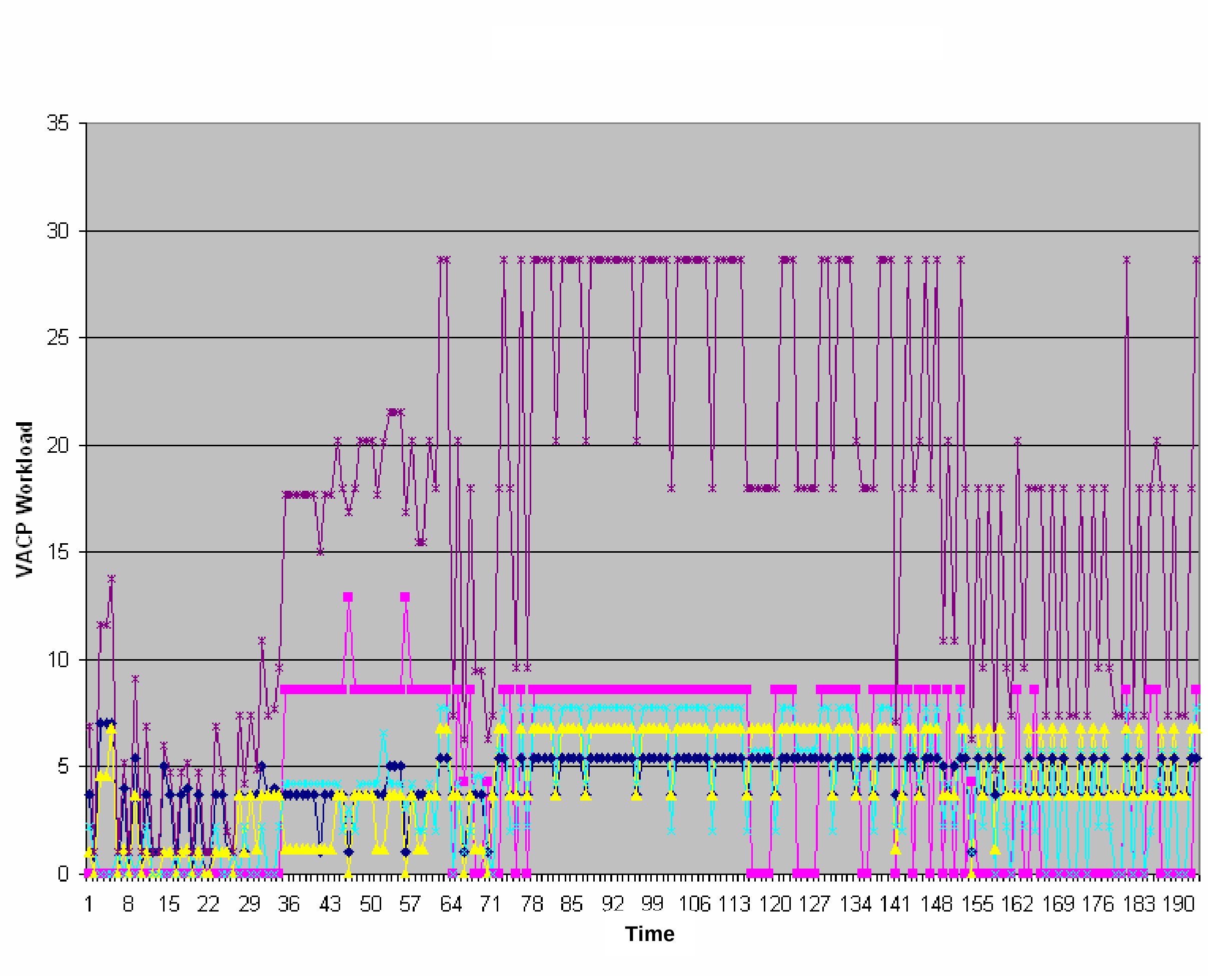
Psychomotor

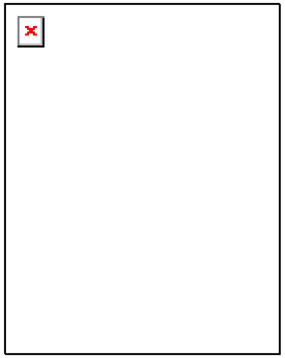
0.0 No psychomotor activity
1.0 Speech
2.2 Discrete actuation
2.6 Continuous adjustment
4.6 Manipulative adjustment
5.8 Discrete adjustment
6.5 Symbolic production (writing)
7.0 Serial discrete manipulation (keyboard entries)

VACP Analysis Process



Typical VACP Analysis Output





VACP Implementation Considerations

- Conceptual/Application Complexity
  - Developing the correct level of task
  - Assigning levels from the resource channel scales
  - Identifying and including *all* the tasks that affect the operator
  - May be time consuming and data intensive
- Positive aspects
  - High validity
  - High diagnosticity
  - Fairly easy to understand

Workload Assessment Methodologies
Subjective Workload Assessment

- The operators are the best judges of how "burdened" they feel

- Different types of rating scales or based on task loading (using a detailed task analysis as the starting point)

- Typically, workload questionnaires are administered immediately (or soon) after a session is completed, but may also be administered:

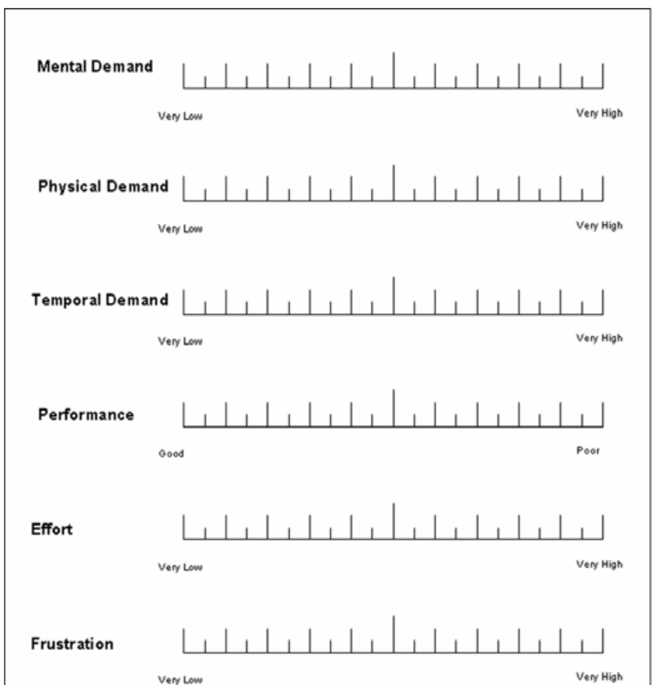
  - In comparison to an arbitrary "baseline" scenario (immediate and comparative)
  - When a set of sessions are completed (delayed and comparative)
  - Retrospective pairwise comparisons (delayed and comparative)

Assessment techniques

- NASA TLX: NASA Task Load Index
- Cooper-Harper
- Modified Cooper-Harper
- Overall Workload
- Workload Profile
- SWAT: Subjective Workload Assessment Techniques
- SWORD: Subjective Workload Dominance

NASA TLX – NASA Task Load Index

- Subjective assessment methodology
- Administered after the task is performed, each session considered individually
- Rate workload along six dimensions from 0 → 100



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 you 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?

NASA TLX Implementation Considerations

- Conceptual/Application Complexity
  - Easy to understand and use
  - Can be implemented manually or in a computer based interface

- Positive aspects
  - High validity
  - High diagnosticity

KEY TASK: Develop appropriate method to assess the impact of operations personnel workload on plant safety

- Assessing operator performance in the overall context of plant safety is done using Human Reliability Analysis (HRA)
- To help support risk-informed decision-making, we chose to use workload assessment results to inform HRA methods
- ATHEANA and SPAR-H are the primary HRA methods currently in use by the NRC
- We demonstrated how workload measures from human performance models using the VACP metrics and empirical data from NASA TLX metrics could be incorporated into ATHEANA and SPAR-H HRA methods
- Many assumptions went into this approach and it has not been applied or validated

ATHEANA – A Technique for Human Event Analysis

- The ATHEANA technique has two components:
  - A retrospective analysis process that seeks to describe the elements of a serious event that has occurred
  - A prospective process that includes a HRA that requires:
    - Estimating the error-forcing context for base case deviations
    - Estimating the conditional likelihood of a human failure event given that context
    - The error forcing context is derived from the description of plant conditions and the performance shaping factors in the plant
- Relies on a multidisciplinary team and seeks out inputs from a wide variety of sources in order to identify error-forcing contexts
- Elements are combined to understand the causes and contributions of human errors

Incorporating workload measures into ATHEANA

- VACP workload and NASA TLX workload data can assist the ATHEANA team in identifying error forcing contexts and changes in Performance Shaping Factors (PSFs) due to changes in operator workload, and in turn, changes in error likelihood
- Points in time, or sustained durations, of particularly high or low operator workload observed during the scenarios or events being analyzed, may indicate the potential for an error forcing context
- Depending on the PSFs under consideration, the changes in workload may also influence the Human Error Probability (HEP) estimates for the PSFs
- Since ATHEANA analyses may vary in scope, depth, and context, the degree to which the outputs of VACP or NASA TLX analyses will be useful will vary as well

SPAR-H - Standardized Plant Analysis Risk-Human Reliability Analysis

- Method/tool to determine the probability that an operator will fail when tasked with performing a Basic Event (BE) (diagnosis or action)
- Analyst estimates the degree to which each of eight Performance Shaping Factors (PSFs) will affect the potential for or likelihood of operator failure (human error probability or HEP)
  - Available Time (AT)
  - Stress/Stressors (Arousal) (SSA)
  - Complexity (C)
  - Experience/Training (ET)
  - Procedures (PR)
  - Ergonomics/HMI (EHMI)
  - Fitness for Duty (F)
  - Work Processes (WP)
- Analyst computes the probability of task failure (P) in the context of where the task will be performed and considering the intertask dependencies in the system

Incorporating workload measures into SPAR-H

- The primary means of integrating workload data into SPAR-H analyses is through the performance shaping factors (PSFs)
- PSF levels clearly reflect workload as it is assessed by VACP and NASA TLX, but the nature of the relationship is not immediately obvious because of conceptual differences in the approaches to assessing workloads and deriving SPAR-H error probabilities
- VACP and NASA TLX workload data can be useful to SPAR-H Analysts in determining and validating PSF levels assigned to tasks

Incorporating VACP measures into SPAR-H

- Individual VACP workload values and/or the simultaneous multiple VACP vale can be informative in deriving the PSF levels applied to a decision or action activity in a SPAR-H analysis

- For example:
  - High overall VACP task demand >>>> Stress/Stressors (Arousal) multiplier > 1
  - Tasks that require multiple VACP resources >>>> Complexity multiplier >1

Incorporating VACP measures into SPAR-H

- Sample of guidance provided

  - **SPAR-H PSF Values and HEP multipliers**

    Stress/Stressors (Arousal) (SSA)
    - Extreme (disruptive) (5)
    - High (2)
    - Nominal (1)
  - **Relationship between PSF and HEP**

    As SSA increases above nominal HEP increases. At extreme SSA levels, catastrophic levels of performance failure can be expected.

Incorporating VACP measures into SPAR-H

- Sample of guidance provided continued
  - **How do the VACP workload demands and Model Time data interact with the PSF to affect HEP? :**
    - High VACP is a stressor, especially if the workload demand is high for more than one resource at the same time or if the demand on any one resource remains high over an extended period of time.
    - The analyst should examine the VACP task workloads for the series of tasks associated with a BE and determine if multiple VACP resources are being demanded simultaneously and if one or more of the resources is "maxed out" during task performance. In these cases, a High or Extreme rating for SSA should be considered.
    - The analyst should also determine if there are extended periods when multiple resources are utilized simultaneously, especially if the demand on each resource is high and there is the potential for interference. In these cases, a High rating should be considered for SSA.
    - If AT is short for a series of high VACP workload tasks, there is an increased likelihood that SSA will be extreme (disruptive) for this BE. This is especially true if PR or EHMI are not good.

Incorporating NASA TLX measures into SPAR-H

- NASA TLX Workload dimensions are grouped according to:
  - Demands imposed by the task (mental, physical and temporal task demands), and
  - How the operator experiences the workload associated with performing the task (experienced effort, perceived performance, and experienced frustration)
- The “fit” between NASA TLX Workload dimension scores, NASA TLX Overall Workload (OW) scores, and SPAR-H PSF ratings is not a direct one-to-one translation
  - For example, the SPAR-H PSFs do not address the full range of possible values – the impacts of things such as low arousal, low complexity, or better procedures is not addressed
- However, NASA TLX ratings provide valuable information to the Analyst in selecting values for PSFs
- We have developed guidance for all of the meaningful combinations (56) of NASA TLX dimensions and SPAR-H PSFs

Incorporating NASA TLX measures into SPAR-H

Sample Guidance

If you have TLX Ratings for Mental Demand (Cognitive/ perceptual demands)	**SPAR-H PSF Values and HEP multipliers**	**Relationship between PSF and HEP**	**How does TLX Mental Demand interact with the PSFs to affect HEP?**
TLX Mental Demand ratings: any number between: Very Low = 0 and Very high = 100	Available Time (AT) Relationships: Diagnosis Task time AT = barely adequate time (2/3 ave time required) (10) AT = sufficient time (1) AT = extra time (.1) AT = expansive time (.1 - .01) Action Task time AT = just enough time (10) AT = some extra time over minimal (1) AT >= 5X time required (.1) AT >= 50X time required (.01)	As AT decreases, the PSF multiplier increases so HEP increases. Increasing mental demands will decrease the relative AT (the ratio of available time/demand will get smaller)	Low TLX Mental (cognitive/perceptual) demands should not alter the effect of AT on HEP, unless mental demands are so low that the operator is bored and loses Situation Awareness. If TLX Mental demands are reported by the operator to be moderate or high the analyst needs to carefully consider whether the AT is adequate. For example, a task that requires interpreting the semantic content of a voice stream could be rated at 60 (moderately demanding) on TLX Mental (cognitive/perceptual) Demand. The SPAR-H analyst may consider that the AT for this task is nominally equal to the time required, but if the sound signal gets distorted or varies in strength, the time the operator needs to perform a task with this Mental Demand may actually become greater than the time allowed, so task P(failure) = 1.

Summary

- VACP and NASA TLX are effective methods for assessing workload
- These methods provide promising and pragmatic approaches that can be integrated with risk-informed decision making processes
- HRA methods such as AHTEANA and SPAR-H provide the link for helping to determine the potential impact of operator workload on plant safety