

5.11 Predicting the Consequences of Workload Management Strategies with Human Performance Modeling

Predicting the Consequences of Workload Management Strategies with Human Performance Modeling

Diane Kuhl Mitchell & Charneta Samms
Army Research Laboratory
Aberdeen Proving Ground, Maryland

diane.k.mitchell@us.army.mil csamms@arl.army.mil

Abstract. Human performance modelers at the US Army Research Laboratory have developed an approach for establishing Soldier high workload that can be used for analyses of proposed system designs. Their technique includes three key components. To implement the approach in an experiment, the researcher would create two experimental conditions: a baseline and a design alternative. Next they would identify a scenario in which the test participants perform all their representative concurrent interactions with the system. This scenario should include any events that would trigger a different set of goals for the human operators. They would collect workload values during both the control and alternative design condition to see if the alternative increased workload and decreased performance. They have successfully implemented this approach for military vehicle designs using the human performance modeling tool, IMPRINT. Although ARL researchers use IMPRINT to implement their approach, it can be applied to any workload analysis. Researchers using other modeling and simulations tools or conducting experiments or field tests can use the same approach.

1.0 INTRODUCTION

As system engineers begin to design a system, it is critical for them to understand how the human operators will interact with the system. This understanding is critical because they are designing the system so the human operators can accomplish specific goals. The humans' ability to accomplish these goals, therefore, determines the effectiveness of the system design. A key component of the human operators' abilities to use the system, in turn, is their mental workload level. Mental workload is a key component because it influences the human operators' performance.

The relationship between human performance and mental workload is often represented as similar to the Yerkes-Dodson (1908) inverted-U relationship as shown in Fig 1. As Fig 1 indicates when mental workload is very low human performance will decline.



Figure 1 Inverted-U relationship between workload and performance (modified from Yerkes & Dodson, 1908).

As workload increases so does human performance. However, at some point workload transitions to a level high enough to overload human mental resources (Wickens, 2008). To manage the high workload, humans employ strategies to reduce workload to manageable levels. These strategies are called workload management strategies (Little, 1993). A strategy, for example, might be to stop an ongoing task, ignore a new task or to perform concurrent tasks sequentially. All of these workload management strategies can result in performance decrements.

For over a decade human performance researchers (Colle & Reid, 2005; Rueb, Vidulich, & Hassoun, 1994; Reid & Colle, 1988; Schlegel, B., Schlegel, R., & Gilliland, 1988; Grier, Wickens, Kaber, Strayer, Boehm-Davis, Traflet, & St. John, 2008) have attempted to refine the inverted-U representation of workload by identifying the point where workload and performance transition from acceptable to unacceptable. They refer to this transition point as the workload redline or threshold (Grier, et al, 2008). Identifying this workload threshold is important. If it could be determined, then human factors researchers could establish a workload level that is considered acceptable for optimum human performance. System engineers, in turn, could use this workload guidance to help ensure their system designs provide effective human performance. Despite the many years of research, there is, however, no consensus among researchers on a workload threshold.

A range of workload threshold values have been proposed by researchers who used the subjective workload assessment tool (SWAT) to estimate

workload. These researchers have proposed SWAT threshold values in the range of 40- 50 (Colle & Reid, 2005; Rueb, Vidulich, & Hassoun, 1994; Reid & Colle, 1988; Schlegel, B., Schlegel, R. & Gilliland, 1988). The SWAT workload range is useful for system engineers conducting system evaluations. In these evaluations human participants can give self-report workload ratings which SWAT requires.

Not all evaluations of system designs, however, include human participants that can give self-report workload ratings. Human performance modeling, for example, is an effective technique for evaluating system designs that includes mental workload evaluation but does not include human participants (Mitchell, 2000). The human operators are simulated in human performance models and, therefore, self-report workload scales, such as, SWAT cannot be used. Using human performance modeling, however, has several advantages over techniques that use human participants.

Human performance modeling is particularly useful early in the system development phase when finding a representative sample of human users of the proposed system can be costly and challenging due to funding constraints. In addition modeling can be used when a representative sample of users is unavailable or only a small sample size of users is available. Finally, it is useful when the design is still a concept and no system mock-ups exist. For human performance modeling techniques that include mental workload prediction as part of the system design evaluation a workload threshold remains critical.

Human performance modelers at the Army Research Laboratory have developed an analytical approach for establishing a workload threshold they can use for evaluation of a proposed system design. Their technique includes three key components. First, they create a scenario containing segments with each segment representing events that change the goals of the operators of the system. Second, they establish a baseline they can use for workload and performance comparisons. Finally, for each of these segments, they select unique workload threshold values for each operator who will operate the system.

In 2009, the ARL modelers implemented this approach in an analysis of the impacts of two conceptual technologies on the workload and performance of a tank crew (Mitchell, in review).

2.0 CASE STUDY

To implement their approach, the ARL modelers used the human performance modeling tool, IMPRINT (Improved Performance Research Integration Tool; <http://www.arl.army.mil/IMPRINT>). IMPRINT is a stochastic task- network modeling tool that provides modelers with the capability to simulate humans performing tasks. The humans simulated for this project were the tank crewmembers. Specifically, the ARL modelers built a model simulating the tasks performed by each crewmember of a baseline tank. Next, they built a model to represent the tasks performed by the tank crewmembers when the vehicle design was enhanced to include a driver's aid and a loader's situation awareness display.

In addition to simulating task performance, IMPRINT also provides modelers with the capability to predict the mental workload associated with the tasks individuals perform (Mitchell, 2000). The ARL modelers used this mental workload option to predict the mental workload of the crewmembers of the baseline tank as well as the enhanced tank. The theoretical basis for the IMPRINT mental workload option is Multiple Resource Theory (MRT) (Wickens, 2008).

According to MRT, the capacity of human mental resources is limited. Therefore, as an individual performs a task, the task makes demands upon these limited mental resources. Furthermore, when an individual performs two or more tasks concurrently, all the concurrent tasks demand some of the individual's mental resources. Because the mental resources have limits, the demands of the concurrent tasks may exceed or overload the individual's resources. The point where the individual's resources are overloaded is the workload threshold. When this threshold is exceeded, the individual implements workload management strategies which cause the individual's performance to decline.

Because the IMPRINT workload capability is based on MRT, its workload predictions are task-based predictions. Changes in the tank crewmembers workload, therefore, are related to changes in the tasks they perform in the baseline versus modified tank. If the technologies in the modified tank reduce crew workload then the IMPRINT workload predictions should be lower for the modified versus baseline tank model runs.

The IMPRINT tool implements MRT by providing modelers with the capability to enter the mental resources required by each task for the human operators of a proposed system. Furthermore, it provides numerical values for estimating the demands of the operators' tasks on their mental workload. IMPRINT provides these numerical values in the form of scales. There are seven scales, one for each resource. The resources represented by the seven scales are visual, auditory, cognitive, fine motor, gross motor, tactile, and speech.

Using the workload scales in IMPRINT, the ARL modelers selected the appropriate values for each of the resources that a tank crewmember used for each task. The IMPRINT software aggregated these workload inputs across all the tasks the crewmember performed every time a new task was started. IMPRINT then provided an overall workload score. This overall workload score is compared to a workload threshold set by the modelers. If the overall workload number exceeded the threshold than a workload management strategy is triggered within the model. Modelers can then see the impact of the crewmember's workload on performance with the system. Because the workload threshold is the key to determining if a workload management strategy is employed, it was critical for the IMPRINT modelers to select an appropriate workload threshold for the tank crew in their analysis.

As the first step in identifying a workload threshold for the tank crew analysis, the ARL modelers selected a scenario to model with IMPRINT. For the performance to be representative of the typical tank crew, the scenario needed to be one in which the crew performed the majority of their common tank crew tasks. These common tank crew tasks are driving, communicating, searching for targets, and engaging targets (Directorate of Training, Doctrine, and Combat Development Field Manual 3-20.15, 2007).

The ARL modelers needed to include common crew tasks in the scenario because they would build the tank crew tasks into the IMPRINT model based on the scenario. It was critical for the IMPRINT workload analysis to be valid that the crew be performing all the tasks the technologies might influence within the model. Furthermore, it was especially important for the ARL modelers to include in the scenario those common crew tasks the crewmembers perform concurrently. The

inclusion of concurrent tasks in the models was important because workload is typically higher, and performance is typically lower, for concurrent tasks than sequential tasks (Just, Carpenter, Keller, Emery, Zajac, & Thulborn, 2001). To meet these scenario characteristics, the modelers selected a movement to contact mission (Directorate of Training, Doctrine, and Combat Development Field Manual 3-20.15, 2007).

After selecting the scenario, the ARL modelers divided the mission into segments that represented changes in the crewmembers' goals. For example, as a movement-to-contact mission begins, the crewmembers' goal is to detect the enemy, whereas, once they detect the enemy their goal shifts to destroying the enemy. As a consequence of the shift in goals between the two segments, the crew performs different tasks. Because the workload predictions in the IMPRINT model are based on task demands, the crews' workload will change along with the tasks. Therefore, if the crewmembers perform a unique set of tasks in one segment than another segment, it is reasonable to assume that their workload will be very different from one segment to another. For example, in mission segments during which the tank is stationary, the driver could engage a target. In contrast, when the vehicle is moving, the driver is driving and would not be engaging targets. The segments the ARL modelers selected to represent diverse sets of crewmember tasks for the movement-to-contact mission were: movement to contact begins, move via checkpoints to the line-of-departure, precision engagement, and move to defensive position.

As they begin the movement-to-contact mission, the goal of the crewmembers is to be ready for the mission. They perform workstation and communications equipment set-up. As they move via checkpoints, their goals shift to searching for potential enemy. They communicate, drive, search for threats, track the battle and do hasty planning. After the enemy is detected, their goals shift again to destroying the threat. They identify, engage, and destroy the threat. Finally, after the enemy is eliminated, their goals shift to avoiding detection by opposing forces. They back-up the vehicle and drive quickly to a defensive position while avoiding enemy detection.

For each of these segments, the ARL modelers set a unique workload threshold. Each threshold was unique because of the variation in tasks, and,

therefore, workload in each mission segment. Furthermore, each crewmember needed a unique workload threshold because the crewmembers performed very different combinations of tasks. For example, the PL does tactical planning, communications monitoring, and supervisory tasks while the driver drives the vehicle. They obtained the threshold values for each crewmember for each segment from an existing baseline IMPRINT tank model. Mitchell (2009) describes this model and the steps the ARL analyst followed in its development in detail.

After developing the mission segments and selecting thresholds, the ARL analysts included the segments in their task-network models. In the IMPRINT task-network models, the ARL modelers represented the sets of tasks the crewmembers performed in each segment of the scenario as functions. Driving, scanning for threats, and communications, for example, would be functions in the model. Furthermore, the task-network model is hierarchical which means functions, at the higher level, can be decomposed into smaller units called tasks. Thus, the ARL modelers decomposed the functions in each segment into tasks. Examples of tasks for the driving function would be maintain speed, adjust steering, monitor forward terrain, etc.

After creating the hierarchical task-network of functions and tasks for each crewmember in each segment of the scenario, the ARL modelers identified the interfaces or equipment the crewmembers used to perform the tasks. IMPRINT provides modelers with the capability to enter the list of interfaces used by the human system operators for each task. Thus the ARL modelers entered the list of interfaces each crewmember used for each task into the baseline tank model. Then, using the IMPRINT workload scales, the modelers estimated the demands that each task and interface combination placed upon the each crewmember's mental resources (visual, auditory, cognitive, gross motor, fine motor, speech, or tactile).

Once the workload data was entered, the ARL modelers ran the baseline tank model multiple times. The multiple runs represented all the possible combinations of functions and tasks that the crewmembers performed during each segment of the mission. Based on these runs the modelers then identified for each crewmember in each mission segment, the combination of tasks that had the highest overall workload value. In addition,

they calculated the average workload across all the runs for each crewmember for each segment. The maximum workload value and average workload value became the workload threshold for that crewmember for that mission segment for the baseline model.

The ARL modelers then modified the baseline model to represent the crewmembers performing the tasks with the two proposed technologies. Specifically, they modified the interfaces used by two of the tank crewmembers, the driver and the loader. Because the interfaces for these two crewmembers were modified from the baseline, the ARL modelers needed to modify the tasks these two crewmembers performed. For example, in the baseline model, a crewmember needed to open the hatch to do a specific task while in the modified model the loader's display enabled the loader to perform with a closed hatch.

When the modified model was complete, the ARL modelers ran it multiple times and calculated the same workload measures as they had for the baseline model. They then compared the two models to see if the crew workload in the modified tank model was higher than the threshold value established from the baseline model. If the workload was the same or lower, they recommended the technologies for further testing. If it was higher than the baseline they recommended evaluating if the potential for overload was mitigated by an increase in performance.

In addition, to the workload comparison, the ARL modelers compared the performance of the crewmembers in the two models. For example, the loader's workload may have remained the same for both models but the technology may have increased his performance by permitting him to do surveillance buttoned-up rather than out-of-the-hatch. Furthermore, a crewmember performing with an open hatch is at a greater risk of injury than with a closed hatch. Greater risk of crew injury, in turn, for represent a great risk to crew survivability and, therefore, the overall movement-to-contact mission.

The overall conclusion of the analysis was that the new technologies did have the potential to increase mission performance while reducing crew workload.

3.0 DISCUSSION

The ARL modelers found the practical approach to establishing the threshold more effective for justifying their recommendations for system design changes than other threshold techniques they had used over the past decade. In earlier efforts, the modelers had used a single overall workload value of 60 (Mitchell, Samms, Henthorn & Wojciechowski, 2003) or 40 (Mitchell, 2005) or 7 (McCracken & Aldrich, 1984) as the workload threshold value. Although these projects changed system requirements (Mitchell & Samms, 2009), and the results were replicated in experiments (Chen, 2009), the selection of a single workload threshold for all crewmembers across a scenario was challenging for the ARL modelers to defend. Because the threshold was difficult to defend, it made it difficult to convince system engineers to change designs. The single threshold value was difficult to defend because the overall workload values from IMPRINT could vary widely between crewmembers due to variations in the functions and tasks they perform. The driver of the tank, for example, might have a maximum workload value of 200, in contrast to the loader who has a maximum workload value of 60. Thus, with a single threshold value of 40, both crewmembers would be overloaded in the baseline but one would have a much higher workload value than the other. In this situation, the crewmember with the workload that exceeded threshold by the most would be most likely to be the focus for system design changes. In comparison, by identifying a threshold for each crewmember, the ARL modelers had more capability to focus attention equally across crewmembers and influence system design changes for each crewmember.

Another challenge confronting the ARL modelers was that the functions, tasks, and workload that a single crewmember performed changed significantly from segment to segment. For example, the highest workload value in an IMPRINT model for a tank driver within a mission might be 200 and the average across the mission might be 100. This high workload is associated with the driving function and tasks. The workload, therefore, would not be representative of the driver's workload when the platform is stationary. During this mission segment, the driver's workload would be 31, a much lower value because the driver is not driving but is scanning for threats. If the mission were not divided into segments this difference in workload would not be apparent

because average workload across the mission would be 115.5 and high workload 200. The practical threshold approach solves this problem by divided the operational scenario into segments representing changes in functions and tasks for crewmembers and the associated workload value changes.

4.0 CONCLUSION

ARL modelers recommend the practical approach to setting a workload threshold be used to evaluate system designs. Although they implemented their approach with the human performance modeling tool, the practical threshold approach can be applied to any workload analysis. Researchers conducting experiments or field tests can use the same approach. To implement the approach in an experiment, the researcher would create two experimental conditions: a baseline and a design alternative. Next they would identify a scenario which includes all the goals of the participants with the system. They would divide this scenario into the segments that represent these goals. They would then have the test participants perform all their representative concurrent interactions with the system in each segment. They would collect workload values during both the control and alternative design condition for each segment and compare workload and performance of the participants in the two conditions. They would then make recommendations based on the workload comparisons.

As a result of this analysis, two enhancements to IMPRINT were recommended. When the ARL modelers analyzed the results across each mission segment, they used the Function Performance report. The Function Performance report provides analysts with detailed information on function duration, accuracy and frequency. This report is generated by looking at all the functions in the model that have started and finished during the model execution but does not report instances where functions are stopped or interrupted. The same is true for the Task Performance report that reports similar information but at the task level. Expanding these reports to include data about function or task stops and interrupts will provide more detailed results to the analyst.

Another recommended enhancement was to allow analyst to choose at what level they would like to define workload thresholds; at the function or mission segments level or at the overall mission

level. Currently, workload thresholds are set per operator over the length of the entire mission. There may be times where different segments of a mission may have different workload thresholds. Implementing this capability in IMPRINT would allow the analyst more flexibility in exploring new workload theories. These enhancements will be considered for implementation in the next IMPRINT development cycle.

5.0 REFERENCES

- [1]. Chen, J. Y. and Joyner, C. T. (2009). *Concurrent Performance of Gunner's and Robotics Operator's Tasks in a Multitasking Environment*. *Military Psychology*, 21(1), 98-113.
- [2]. Colle, H. A. & Reid, G. B. (2005). *Estimating a mental workload redline in a simulated air-to-ground combat mission*. *The International Journal of Aviation Psychology*, 15 (4), 303-319.
- [3]. Directorate of Training, Doctrine, and Combat Development. (2007) *Tank Platoon (FM 3-20.15)*, 204 1st Cavalry Regiment Rd, Ste. 207, U.S. Army Armor Center, Fort Knox, KY 40121-5123, 2007.
- [4]. Grier, R., Wickens, C., Kaber, D., Strayer, D., Boehm-Davis, D., Trafton, J. G., & St. John, M. (2008). *The red-line of workload: theory, research, and design*. Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting, New York, pp. 1204-1208.
- [5]. Just, M. A., Carpenter, P. A., Keller, T. A., Emery, L. Zajac, H. & Thulborn, K. R. (2001). *Interdependence of non-overlapping cortical systems in dual cognitive tasks*. *NeuroImage*, 14, 417-426.
- [6]. Little, R. (1993). *Final report for the crew reduction in armored vehicles ergonomic (CRAVES) study (CDRL No. A006)*. Human Engineering Laboratory, Aberdeen Proving Ground, MD.
- [7]. McCracken, J. H. & Aldrich, T. B. (1984). *Analyses of Selected LHX Mission Functions: Implications for Operator Workload and System Automation Goals, (ARI479-024-84)*. Army Research Institute, Fort Rucker, AL.
- [8]. Mitchell, D. K. (in press). *Abrams V2 SEP Crew Workload Analysis: Impacts of Two Proposed Technologies*. U.S. Army Research Laboratory, Aberdeen Proving Ground, MD.
- [9]. Mitchell, D. K. (2009). *Workload Analysis of the Crew of the Abrams V2 SEP: Phase I Baseline IMPRINT Model (ARL-TR-5028)*. Army Research Laboratory, Aberdeen Proving Ground, MD.
- [10]. Mitchell, D. K. (2005). *Soldier Workload Analysis of the MCS Platoon's Use of Unmanned Assets (ARL-TR-3476)*. U.S. Army Research Laboratory, Aberdeen Proving Ground, MD.
- [11]. Mitchell, D. K. (2000). *Mental Workload and ARL Workload Modeling Tools (ARL-TN-161)*. U.S. Army Research Laboratory, Aberdeen Proving Ground, MD.
- [12]. Mitchell, D. K. & Samms, C. L. (2009). *Workload Warriors: Lessons Learned from a Decade of Mental Workload Prediction Using Human Performance Modeling*, Human Factors and Ergonomics Society 53rd Annual Meeting Proceedings, 53(12), 819-823.
- [13]. Mitchell, D. K., Samms, C. L., Henthorn, T., & Wojciechowski, J. (September, 2003). *Trade Study: A Two- Versus Three-Soldier Crew for the Mounted Combat System (MCS) and Other Future Combat System Platforms (ARL-TR-3026)*. U.S. Army Research Laboratory, Aberdeen Proving Ground, MD.
- [14]. Reid, G. B. & Colle H. A. (1988). *Critical SWAT values for predicting operator overload*. Proceedings of the Human Factors Society – 32nd Annual Meeting. Santa Monica, CA: Human Factors Society, 1414-1418.
- [15]. Rueb, J.D., Vidulich, M.A. & Hassoun, J.A. (1994). *Use of workload redlines: A KC-135 crew-reduction evaluation*. *International Journal of Aviation Psychology*, 4, 47-64.
- [16]. Schlegel, B., Schlegel, R. E., & Gilliland, K. (1988). *Determining excessive mental workload with the subjective workload assessment technique*. Proceedings of the 10th Congress of the International Ergonomics Association (Vol. II), Sydney Australia, 475-477.
- [17]. Wickens, C. (2008). *Multiple Resources and Mental Workload*. *The Journal of the Human Factors and Ergonomics Society*, 50(3), 449-455.
- [18]. Yerkes, R. M. & Dodson, J. D. (1908). *The relation of strength of stimulus to rapidity of habit-formation*. *Journal of Comparative Neurology and Psychology*, 18, 459-482.

Predicting the Consequences of Workload Management Strategies with Human Performance Modeling

Charneta Samms
Diane Mitchell
U.S. Army Research Laboratory

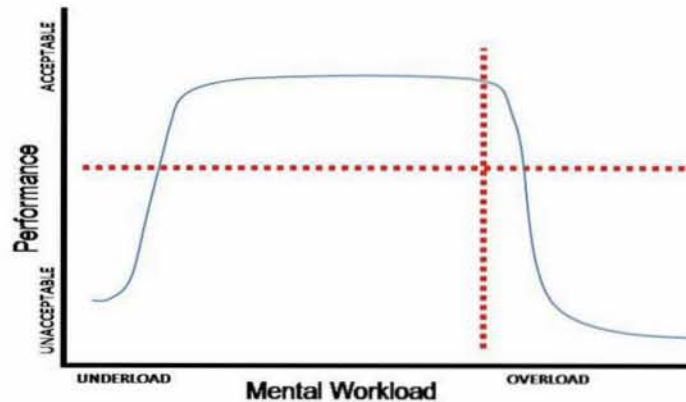


October 13-15, 2010

Agenda

- Mental workload and human performance
- Human Performance Modeling (HPM)
- Improved Performance Research Integration Tool (IMPRINT)
- Analysis Approach
- Case Study
- Conclusions

Mental Workload and Human Performance



Inverted-U relationship between workload and performance

Modified from Yerkes, R. M. & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation, *Journal of Comparative Neurology and Psychology*, 18, 459-482.

Importance of Workload

- Indicator of problem areas within system design
- Peaks and valleys of workload indicate times when human performance may suffer, e.g.:
 - Sustained low workload (underload) leads to boredom, loss of situation awareness, and reduced alertness.
 - Sustained high workload (overload) leads to fatigue.
 - Workload peaks lead to dropped tasks, increased task time, cognitive tunneling, and increased errors.
- Reduces crew performance, system performance, and contribute to mission failure

OBJECTIVE: Achieve evenly distributed, manageable workload. Avoid both overload and underload.

Why Human Performance Modeling (HPM)?

Concept System

Many Variables

Field Study Not Feasible

Too Dangerous

Model – Test – Model

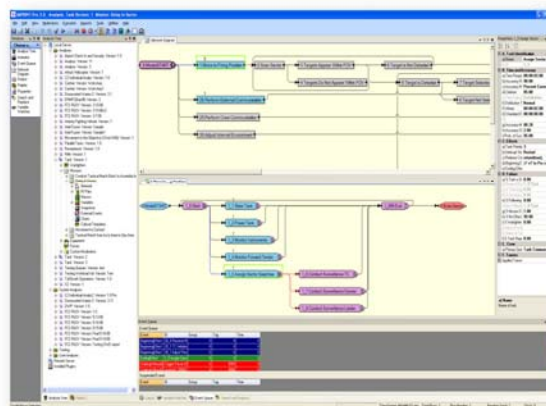
System Performance $\cong f(\text{human performance})$

5

Improved Performance Research Integration Tool



334 users supporting Army, Navy, Air Force, Marines, NASA, Department of Homeland Security (DHS), Department of Transportation (DoT), Joint and other organizations across the country



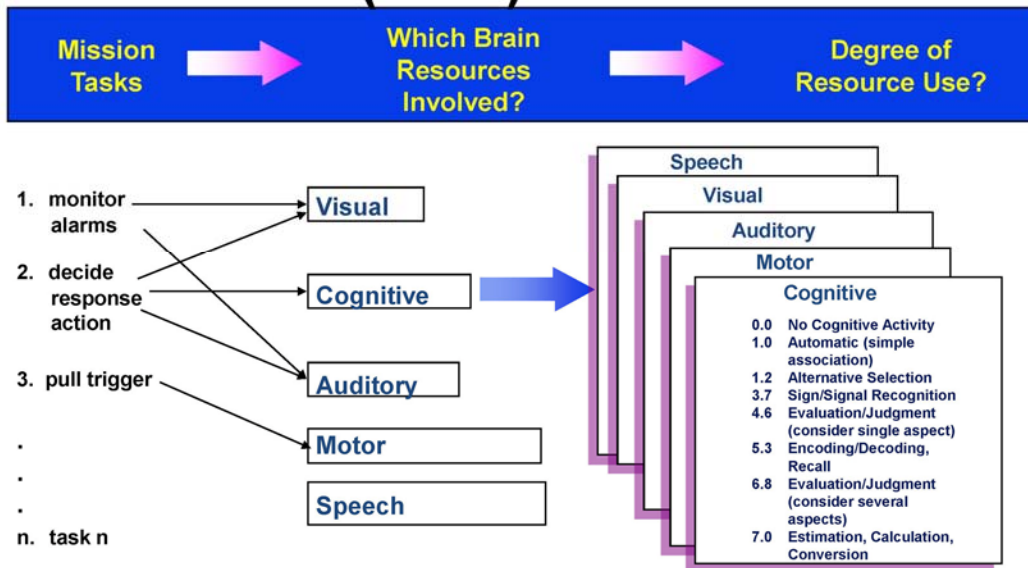
<http://www.arl.army.mil/IMPRINT>

IMPRINT can be used to

- Set realistic system requirements
- Identify future manpower & personnel constraints
- Evaluate operator & crew workload
- Test alternate system-crew function allocations
- Assess required maintenance man-hours
- Assess performance during extreme conditions
- Examine performance as a function of personnel characteristics and training frequency & recency
- Identify areas to focus test and evaluation resources
- Quantify human system integration risks in mission performance terms to support milestone review
- Represent humans in federated simulations

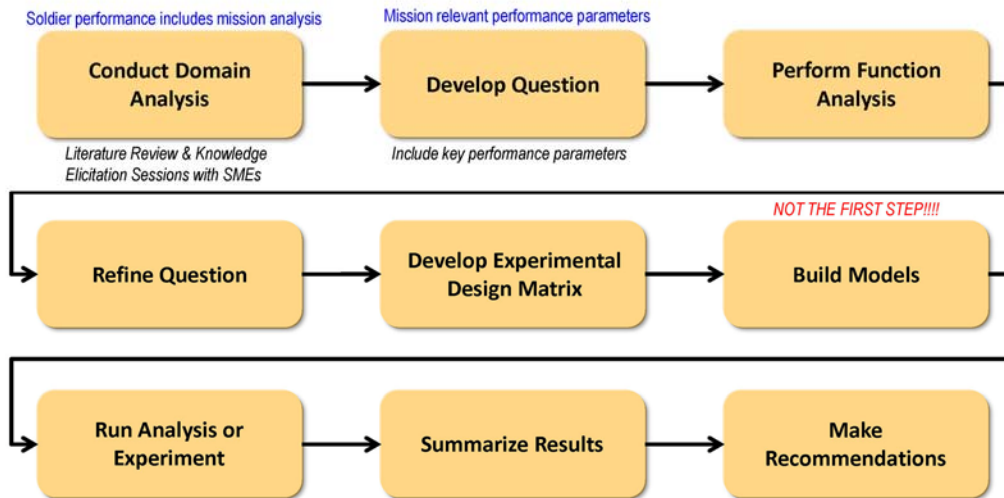
IMPRINT is a trade-off analysis tool

Multiple Resource Theory (MRT) in IMPRINT

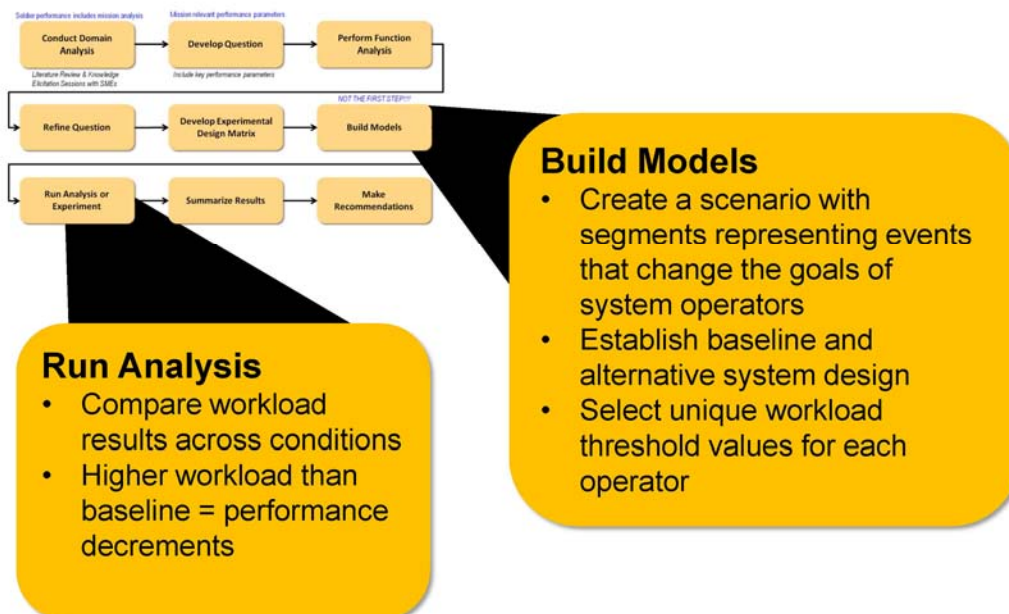


Analysis Approach

Quantify influence of human operator performance on system/mission performance



Executing the Approach

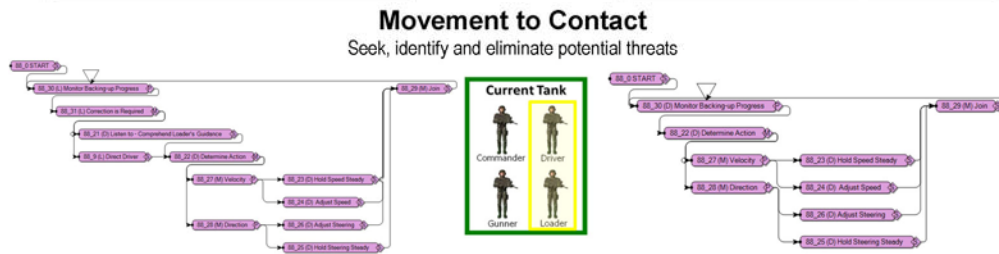


Case Study

Examine impact of two conceptual technologies on workload and system performance

BASELINE model
Without technologies

ALTERNATIVE model
With technologies



New technologies have potential to increase mission performance while reducing crew workload

Mitchell, D. K. (in press). Abrams V2 SEP Crew Workload Analysis: Impacts of Two Proposed Technologies. U.S. Army Research Laboratory, Aberdeen Proving Ground, MD.

Summary

- Use analysis approach to setting workload thresholds in HPM or experimentation
 - Develop overarching scenario
 - Set up at least two conditions; e.g. baseline & alternative
 - Compare workload levels
 - Make recommendations based on workload comparisons
- Potential enhancements for IMPRINT
 - Expansion of function & task performance reports
 - Function level workload thresholds