# Establishing Workload Manipulations Utilizing a Simulated Environment

Julian Abich IV, Lauren Reinerman-Jones, and Grant Taylor

University of Central Florida (UCF), Institute for Simulation & Training (IST),
Applied Cognition in Virtual Immersive Training Environments Laboratory (ACTIVE Lab)
{jabich,lreinerm,gtaylor}@ist.ucf.edu

**Abstract.** Research seeking to improve the measurement of workload requires the use of established task load manipulations to impose varying levels of demand on human operators**.** The present study sought to establish task load manipulations for research utilizing realistically complex task environments that elicit distinct levels of workload (i.e. low, medium, and high). A repeated measures design was used to test the effects of various demand manipulations on performance and subjective workload ratings using the NASA-Task Load Index (TLX) and Instantaneous Self-Assessment technique (ISA). This experiment successfully identified task demand manipulations that can be used to investigate operator workload within realistically complex environments. Results revealed that the event rate manipulations had the most consistent impact on performance and subjective workload ratings in both tasks, with each eliciting distinct levels of workload.

**Keywords:** Workload, simulated environments, complex systems, signal detection, change blindness.

## 1 Introduction

Despite over 50 years of research related to workload, there is not yet an agreed upon definition that captures the complexity of this construct in its entirety (Knowles, 1963; Taylor, 2012). Early concepts of workload focused on physical load (Meshkati, Hancock, Rahimi, & Dawes, 1995), but later emphasized the cognitive components of the construct. Subsequent efforts focused on defining workload, establishing a workload assessment procedure, and finding the relationship among measurements (Johannsen, 1979), but have yet to establish a single, comprehensive operational definition. A common theme across all proposed workload theories is a dynamic interaction between the task and the operator (Veltman & Gaillard, 1996) with the operator having a limited capacity of resources to allocate to the demands of a task (Wickens & Hollands, 2000). Some theories argue that resources come from a single pool of energy (Kahneman, 1973; Moray, 1967), while others argue that multiple pools or multiple levels exist (Wickens, 2002). These theories have been tested using a range of tasks and tasking environments with workload assessed by various measures. However, establishing distinct levels of workload for tasks have not been

R. Shumaker (Ed.): VAMR/HCII 2013, Part II, LNCS 8022, pp. 211–220, 2013.
© Springer-Verlag Berlin Heidelberg 2013

identified to directly compare the universality or sensitivity of the various workload measures. To better understand the role of workload within complex task environments, similar demand manipulations that elicit distinct levels of workload have to be determined. The present study sought to develop these manipulations for two theoretically different tasks related to the military intelligence, surveillance, and reconnaissance (ISR) domain within a simulated unmanned vehicle environment for use in future experimentation.

ISR tasks focus on acquiring and processing information on which to base decisions. Two popularly ascribed theories that form the foundation for tasks within the ISR domain are signal detection theory (SDT) and change blindness (CB). SDT is the foundation for threat detection in ISR missions. SDT states that nearly all decisions are made in the presence of uncertainty (Green & Swets, 1966; Heeger, 1997). This uncertainty results from the presence of noise. Noise is either internal (referring to perceptual processing and/or neural activity), external (referring to environmental sources), or a combination of both. Changes in noise increases or decreases workload, which influences decisions made regarding detection of relevant information. Determining the decision-making criteria to reduce unwanted outcomes or experienced workload (See, Howe, Warm, & Dember, 1995) would be highly advantageous in high-risk environments such as impending threat.

Another task required in ISR missions is monitoring changes to operational area and this type of task is founded on theories of CB. CB refers to the observer's failure to notice change in a visual scene (Rensink, 1997), even if it is large and normally easily noticed (Simon & Rensink, 2005). The visual processing involved in first noticing a change requires detection (whether a change occurred), identification (what kind of change occurred) and localization (where the change occurred; Rensink, 2002). Some theories argue that the inability to detect a change might be due to the operator's limited attentional resources (Kahneman, 1973), limited processing of attended information (Rensink, Regan, & Clark, 1997), failure to compare pre- and post-representations of processed visual information (Hollingworth & Henderson, 2002), or that post-representations overwrite pre-representations (Rensink, Regan, & Clark, 1997). Attentional resources are strongly linked to workload level and therefore, varying levels of workload affects performance on CB tasks. This is particularly important to ISR missions because the inability to notice a change might lead to misdiagnosis of activity occurring in the environment, which could result in decreased safety.

Simulated environments are used to experimentally test the types of theoretically - based tasks described above within complex systems. These environments offer a host of advantages in comparison to real-world testing, such as reduced costs associated with developing, running, and maintaining these systems, consistency and control of variable manipulations, logging capabilities for real-time and post-hoc analysis, and increased safety for consequences resulting from operator error. Simulated environments provide the flexibility to test varying levels of task load manipulations that otherwise would not be possible within a real-world system, ideally resulting in recommendations for design of complex systems and improve operator assessment and training.

## 2 Experiment Overview

Using a simulated unmanned vehicle environment, the aim for the present study was to establish task load manipulations for two theoretically different tasks that induce distinct levels of workload measured by both subjective assessments and performance. The intention was to identify at least one type of manipulation for each task and will determined by the three different types of workload measures. Measures included post-task and online, and subjective and objective assessments. The established manipulations will be implemented in future studies with the goal of observing operator state using physiological measures with the purpose of identifying a universal and comprehensive measure that assesses workload across tasks, domains, and other workload measures.

## 3 METHODS

### 3.1 Participants

Fifty-six (34 males, 22 females) volunteers from several universities located in central Florida participated in the experiment with a mean age of 20.6 (*SD* = 3.4).

### 3.2 Materials

**Questionnaires**

*NASA-Task Load Index (TLX).* The TLX is a multi-dimensional questionnaire used to assess perceived workload (Hart & Staveland, 1988, Hart, 2006). It consists of six subscales of workload (mental demand, physical demand, temporal demand, effort, frustration, and perceived performance), each rated on a 100-point scale with five-point increments and 100 being high workload. The average score of the six-subscales provides a separate measure of global workload. The TLX was administered post-task using a customized computer program to automatically activate a visual prompt containing the questionnaire at multiple designated locations (the end of each trial block) throughout each scenario.

*Instantaneous Self-Assessment (ISA).* The ISA is a single measure used to assess immediate subjective workload during the performance of a task (Tattersall & Foord, 1996). The scale uses a five-point rating scale with 5 being high workload and was administered at multiple designated locations (at 75% of each trial block completion) using a customized computer program to automatically activate an audio prompt containing the phrase “please rate your workload.” Participants responded verbally with their rating.

**Apparatus**

*Simulation.* The Mixed Initiative eXperimental (MIX) testbed (Reinerman-Jones, Barber, Lackey, & Nicholson, 2010) was utilized to simulate an operator control unit

(OCU) for an unmanned ground vehicle (UGV). The environment the UGV maneuvered through was a generic Middle Eastern town infiltrated by enemy threats. The UGV was fully autonomous and drove itself along a preplanned route while participants identified static enemy targets within the environment. The simulation was presented using a standard desktop computer (3.2GHz, Intel Core i7 processor) with a 22” (16:10 aspect ratio) monitor. Responses were collected using the left mouse button and verbal responses to the ISA measure were collected using a standard external desktop computer microphone.

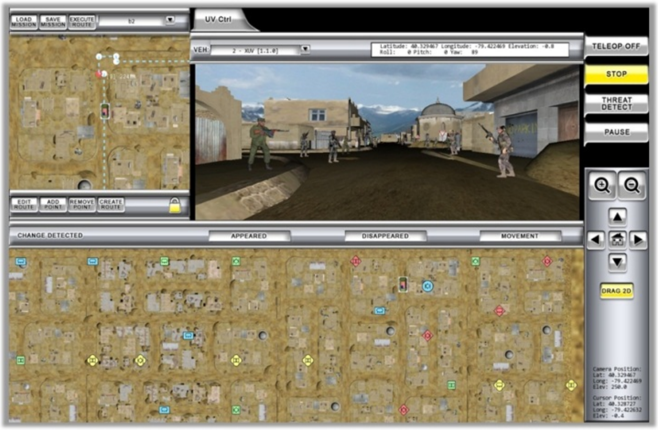


**Fig. 1.** Screenshot of the MIX testbed

**Experimental Tasks.** Participants performed two tasks both independently. One task was the threat detection task based on SDT and the other was the change detection task based on CB theory.

*Threat Detection (TD) Task.* Participants monitored a video feed of the forward perspective of the UGV while it traveled along a pre-planned route and reported any potential threats present in the environment. Four categories of people were present: friendly soldiers, friendly civilians, enemy soldiers, and insurgents (armed civilians). A threat was classified as enemy soldiers and/or insurgents. They were reported by using the computer mouse to left-click a “threat detect” button located along the top right of the OCU and then to left-click on the threat within the UGV video feed. Performance of was rated as percentage of targets correctly identified.

The TD task manipulated both event rate and signal/noise ratio (threat probability). These levels were derived from a meta-analysis of the sensitivity decrement in vigilance (See, Howe, Warm, & Dember, 1995; Table 1). Event rate and threat probability were combined to form five total conditions. Each event rate was presented with a medium threat probability and each threat probability was paired with the medium event rate.

*Change Detection (CD) Task.* Participants monitored an aerial map positioned at the bottom of the OCU that displayed the location of various entities. Entities were represented by icons borrowed from the DoD, but the associated meanings were not tied to other events. On average 24 icons were present and randomly displayed across the defined area. Icons exhibited three types of changes: appear, disappear and movement. Three change detection buttons labeled after each type of change were located above the aerial map. Icon changes identified were reported using the mouse to left-click on the appropriate change detection button. Performance was rated as percentage of changes correctly detected.

The CD task manipulated both event rate and signal saliency. These levels were derived from previous research (Tollner, 2006; Taylor, Reinerman-Jones, Cosenzo, & Nicholson, 2010; Table 1). Event rate and signal saliency were combined to form five total conditions. Each event rate was presented with a medium signal saliency and each signal saliency was paired with the medium event rate.

**Table 1.** Levels of manipulations for both tasks

| Task Demand | Low | Medium | High |
|---|---|---|---|
| | | Threat Detection | |
| Threat Probability | 1/15 | 2/15 | 4/15 |
| Event Rate | 15/min | 30/min | 60/min |
| | | Change Detection | |
| Signal Saliency | 4 icons | 2 icons | 1 icon |
| Event Rate | 6/min | 12/min | 24/min |

## 3.3 Procedure

Participants were trained on the tasks followed by a brief practice session. Each participant completed two scenarios. The two scenarios consisted of the: CD task with the five conditions of workload and TD task with the five conditions of workload (Table 2). Scenario and workload condition order were counterbalanced for each participant. Each block was six minutes, totaling 30 minutes per scenario, with the entire experiment lasting roughly two hours.

**Table 2.** Example of full experiment run

| Tasks | Change detection | Threat Detection |
|---|---|---|
| Manipulations | *Signal saliency/Event Rate ratios:* 2:6, 2:12, 2:24, 1:12, 4:12 | *Threat ratios:* 2:30, 4:30, 8:30, 2:15, 8:60 |

# 4 Results

Evaluation of each manipulation on subjective and performance data was conducted using a series of repeated measure ANOVAs. Bonferroni and Greenhouse-Geisser corrections and pairwise deletions were applied where appropriate. Due to a logging error, the sample size for threat detection performance was reduced to 29.

## 4.1 Threat Detection

Results revealed that the event rate manipulation had a significant main effect on all dependent variables, Mental Demand, $F(1.562, 82.781) = 19.944$, $p < .001$, $\eta^2 = .273$, Physical Demand, $F(2, 106) = 6.354$, $p = .002$, $\eta^2 = .107$, Temporal Demand, $F(1.436, 76.118) = 18.484$, $p < .001$, $\eta^2 = .259$, Effort, $F(1.572, 83.305) = 16.147$, $p < .001$, $\eta^2 = .234$, Performance, $F(1.798, 95.307) = 13.512$, $p < .001$, $\eta^2 = .203$, Global, $F(1.639, 78.654) = 17.065$, $p < .001$, $\eta^2 = .262$, ISA, $F(2, 98) = 48.926$, $p < .001$, $\eta^2 = .500$, and percent correct detected, $F(1.549, 43.371) = 34.305$, $p < .001$, $\eta^2 = .551$, with the exception of Frustration, $F(2, 106) = 4.631$, $p = .012$, $\eta^2 = .080$ (Table 3). Means and standard deviations (in parentheses) for tables 3-6 are provided for each level of demand. Means designated with subscripts are significantly different from means with equivalent subscripts ($p < .0167$ in each case).

**Table 3.** Results for event rate manipulation on threat detection task

| | Level of Demand | | |
|---|---|---|---|
| Variables | Low | Medium | High |
| Mental Demand | $18.056_a$ (20.50) | $23.148_a$ (21.57) | $30.648_a$ (24.32) |
| Physical Demand | $8.80_a$ (14.37) | 10.28 (15.06) | $13.24_a$ (18.54) |
| Temporal Demand | $15.00_a$ (17.88) | $20.65_a$ (20.31) | $27.59_a$ (25.31) |
| Effort | $19.81_a$ (21.50) | $24.63_b$ (23.59) | $31.20_{ab}$ (25.23) |
| Performance | $7.41_a$ (9.85) | $9.26_b$ (9.24) | $13.70_{ab}$ (13.07) |
| Global | $27.13_a$ (11.91) | $30.32_a$ (13.39) | $34.27_a$ (15.34) |
| ISA | $1.42_a$ (.54) | $1.82_a$ (.60) | $2.26_a$ (.75) |
| Percent Detect | $58.27_a$ (16.73) | $68.30_a$ (7.81) | $80.94_a$ (6.63) |

The threat probability manipulation had a significant main effect on Temporal Demand, $F(1.478, 78.34) = 6.83$, $p = .005$, $\eta^2 = .114$, Global, $F(1.651, 79.256) = 6.216$, $p = .005$, $\eta^2 = .115$, ISA, $F(1.567, 76.798) = 22.424$, $p < .001$, $\eta^2 = .314$, and percent correct detected, $F(1.660, 46.485) = 10.342$, $p < .001$, $\eta^2 = .270$, with the exception of Mental Demand, $F(1.704, 90.302) = 5.384$, $p = .009$, $\eta^2 = .092$, Effort, $F(2, 106) = 3.343$, $p = .039$, $\eta^2 = .059$, Physical Demand, $F(2, 106) = 2.152$, $p = .121$, $\eta^2 = .039$, Frustration $F(1.518, 80.45) = 1.103$, $p = .323$, $\eta^2 = .020$, and Performance, $F(2, 106) = 3.074$, $p = .05$, $\eta^2 = .055$ (Table 4).

**Table 4.** Results for threat probability manipulation on threat detection task

| | Level of Demand | | |
|---|---|---|---|
| Variables | Low | Medium | High |
| Temporal Demand | $17.22_a$ (19.00) | 20.65 (20.31) | $24.44_a$ (23.77) |
| Global | $28.15_a$ (12.66) | 30.32 (13.39) | $32.30_a$ (14.21) |
| ISA | $1.52_a$ (.61) | $1.82_a$ (.60) | $2.12_a$ (.77) |
| Percent Detect | $68.10_a$ (17.35) | $68.30_b$ (7.81) | $78.74_{ab}$ (8.32) |

## 4.2 Change Detection

Results revealed that the event rate manipulation had a significant main effect on Mental Demand, $F(2, 98) = 18.460$, $p < .001$, $\eta^2 = .274$, Physical Demand, $F(2, 98) = 6.775$, $p = .002$, $\eta^2 = .121$, Temporal Demand, $F(1.754, 85.502) = 24.958$, $p < .001$, $\eta^2 = .337$, Effort, $F(1.627, 79.723) = 15.111$, $p < .001$, $\eta^2 = .236$, Frustration, $F(2, 98) = 10.227$, $p < .001$, $\eta^2 = .173$, Performance, $F(2, 98) = 8.849$, $p < .001$, $\eta^2 = .153$, Global, $F(1.732, 84.862) = 20.007$, $p < .001$, $\eta^2 = .290$, ISA, $F(2, 96) = 27.466$, $p < .001$, $\eta^2 = .364$, and percent correct detected, $F(2, 94) = 116.856$, $p < .001$, $\eta^2 = .713$ (Table 5).

The signal saliency manipulation only had a significant main effect on percent correct detected, $F(2, 94) = 86.472$, $p < .001$, $\eta^2 = .648$. Signal saliency did not have a significant main effect on Mental Demand, $F(1.476, 72.325) = 1.779$, $p = .184$, $\eta^2 = .035$, Physical Demand, $F(2, 98) = .826$, $p = .441$, $\eta^2 = .017$, Temporal Demand, $F(1.462, 71.638) = 2.189$, $p = .133$, $\eta^2 = .043$, Effort, $F(1.622, 79.472) = 2.114$, $p = .137$, $\eta^2 = .041$, Frustration, $F(1.536, 75.271) = .061$, $p = .899$, $\eta^2 = .001$, Performance, $F(2, 98) = .783$, $p = .460$, $\eta^2 = .016$, Global, $F(1.366, 66.910) = 1.602$, $p = .213$, $\eta^2 = .032$, and ISA, $F(1.532, 73.528) = 1.582$, $p = .215$, $\eta^2 = .032$ (Table 6).

**Table 5.** Results for event rate manipulation on change detection task

| Variables | Level of Demand | | |
|---|---|---|---|
| | Low | Medium | High |
| Mental Demand | $47.90_a$ (26.75) | $56.70_a$ (27.53) | $63.40_a$ (25.30) |
| Physical Demand | $19.00_a$ (17.87) | 22.70 (21.93) | $27.20_a$ (25.86) |
| Temporal Demand | $41.00_a$ (27.31) | $47.70_a$ (28.56) | $58.60_a$ (26.97) |
| Effort | $49.40_a$ (25.85) | $53.60_b$ (26.03) | $62.50_{ab}$ (25.10) |
| Frustration | $31.20_a$ (25.73) | $33.00_b$ (25.50) | $43.30_{ab}$ (28.24) |
| Performance | $34.50_a$ (20.08) | $35.90_b$ (22.08) | $43.20_{ab}$ (22.20) |
| Global | $42.33_a$ (15.04) | $46.30_a$ (16.11) | $51.97_a$ (15.99) |
| ISA | $2.47_a$ (.89) | $2.77_a$ (.94) | $3.39_a$ (.86) |
| Percent Detect | $61.16_a$ (16.82) | $59.41_b$ (15.74) | $38.93_{ab}$ (11.13) |

**Table 6.** Results for signal saliency manipulation on change detection task

| Variables | Level of Demand | | |
|---|---|---|---|
| | Low | Medium | High |
| Percent Detect | $70.06_a$ (15.87) | $59.41_a$ (15.74) | $48.55_a$ (11.20) |

## 5 Discussion

The goal for the present study was achieved. Event rate for both threat detection and change detection elicited distinct levels of workload as shown by the TLX, ISA, and performance. These conditions will be used to investigate physiological responses to distinct levels of workload for two different theoretically driven tasks. That will enable improved adaptive trainers for complex systems, enable direct human-robot implicit communication, and better objective workload assessments.

On the surface, threat probability revealed some distinction between low, medium, and high workload levels. Global TLX showed a trend for these distinct level differences, but not all were significant as indicated by post-hoc comparisons. The same was the case for performance. Furthermore, the TLX sub-scale of temporal demand indicated differences in workload for the low and high levels of signal probability. Thus, the driving factor of this manipulation appears to be time, which is likely due to the amount of time participants felt they had to click on the increased threats in the environment. ISA results did indicate distinct levels of low, medium, and high workload. The inconsistency between the measures indicates that further investigation into signal probability is needed. A future study should look at different probabilities of the signal to noise ratio with perhaps greater increments in the probabilities. Additionally, the type of event should be further investigated. The present study used enemy soldiers and insurgents as signals. It is possible that this manipulation would provide distinct levels of workload if objects like Improvised Explosive Devices (IEDs) were used.

Signal saliency for CD yielded inconsistent results across the three measures of workload. Thus, the manipulation of saliency is not as clearly understood as that of event rate and might not be a manipulation of workload at all, but perhaps drawing on different cognitive capacities. Alternatively, the popularly used subjective measures of workload might not be sensitive to saliency changes that actually do increase or decrease operator workload. Further research should investigate additional subjective workload measures for signal saliency.

The present study illustrates the importance of systematically investigating manipulation choices for experiments and careful consideration of the generalizability of a type of manipulation to various tasks driven by different theories and cognitive processing requirements. It is also important to understand the strengths and limitations of measures for a given phenomenon and not just blindly accept the most popularly used measures.

**Acknowledgements.** This work was supported by the US Army RDECOM (W91CRB08D00150068). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of RDECOM or the US Government. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

## References

1. Green, D.M., Swets, J.A.: Signal detection theory and psychophysics. Wiley, New York (1966)
2. Hart, S.G.: Nasa-Task Load Index (NASA-TLX); 20 Years Later. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 50(9), 904–908 (2006)
3. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, pp. 139–184. Elsevier Science Publishers, Holland (1988)
4. Heeger, D.: Signal Detection Theory. Standford University, Department of Psychology (1997)
5. Hollingsworth, A., Henderson, J.M.: Accurate visual memory for previously attended objects in natural scenes. Journal of Experimental Psychology: Human Perception and Performance, 113–136 (2002)
6. Johannsen, G.: Workload and workload measurement. In: Moray, N. (ed.) Mental Workload, pp. 3–12. Plenum Press, New York (1979)
7. Kahneman, D.: Attention and effort. Prentice Hall, Englewood Cliffs (1973)

8. Knowles, W.B.: Operator loading tasks. Human Factors 5(2), 155–161 (1963)
9. Meshkati, N., Hancock, P., Rahimi, M., Dawes, S.M.: Techniques in mental workload assessment. In: Wilson, J.R., Corlett, E.N. (eds.) Evaluation of Human Work: A Practical Ergonomics Methodology, 2nd edn., pp. 749–782. Taylor & Francis, Bristol (1995)
10. Moray, N.: Where is capacity limited? A Survey and A Model Acta Psychologica 27, 84–92 (1967)
11. Reinerman-Jones, L., Barber, D., Lackey, S., Nicholson, D.: Developing methods for utilizing physiological measures. In: Applied Human Factors and Ergonomics, Miami, FL (2010)
12. Rensink, R.A.: How much of a scene is seen? The role of attention in scene perception. In: Investigative Ophthalmology and Visual Science, vol. 38 (1997)
13. Rensink, R.A.: Change detection. Annual Review of Psychology 53, 245–277 (2002)
14. See, J.E., Howe, S.R., Warm, J.S., Dember, W.N.: Meta-analysis of the sensitivity decrement in vigilance. Psychological Bulletin 117(2), 230–249 (1995)
15. Simons, D.J., Rensink, R.A.: Change blindness: Past, present, and future. Trends in Cognitive Sciences 9(1), 16–20 (2005)
16. Rensink, R.A., Regan, J.K.O., Clark, J.J.: To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science 8(5), 368–373 (1997)*
17. Taylor, G.S.: Comparing types of adaptive automation within a multi-tasking environment. University of Central Florida, Orlando (2012) (Dissertation)
18. Tattersall, A.J., Foord, P.S.: An experimental evaluation of instantaneous self-assessment as a measure of workload. Ergonomics 39(5), 740–748 (1996)
19. Tollner, A.M.: Individual and Team Susceptibility to Change Blindness. University of Cincinnati, Cincinnati (2006) (Dissertation)
20. Veltman, J.A., Gaillard, A.W.K.: Physiological indices of workload in a simulated flight task. Biological Psychology 42, 323–342 (1996)
21. Wickens, C.D., Hollands, J.G.: Engineering psy-chology and human performance, 3rd edn. Prentice Hall, Upper Saddle River (2000)
22. Wickens, C.: Multiple resources and performance prediction. Theoretical Issues in Ergonomic Science 3(2), 159–177 (2002)