

# Contrasting Instructional Strategies Suited to a Detection Task: Examining Differences in Subjective Workload

Crystal Maraj<sup>1</sup>(✉), Jonathan Hurter<sup>1</sup>, William Aubrey<sup>1</sup>, Elizabeth Wolfe<sup>1</sup>,  
and Irwin Hudson<sup>2</sup>

<sup>1</sup> Institute for Simulation and Training, University of Central Florida, Orlando, FL, USA  
{cmaraj, jhurter, waubrey, ewolfe}@ist.ucf.edu

<sup>2</sup> U.S. Army Research Laboratory – Human Research and Engineering Directorate  
Advanced Training and Simulation Division, Orlando, FL, USA  
irwin.l.hudson.civ@mail.mil

**Abstract.** Soldiers benefit from the ability to detect threats conveyed via human kinesic cues, or non-verbal body movements. Simulation-Based Training (SBT) supplies an avenue to improve kinesic cue detection performance. Instructional strategies in SBT are designed to enhance performance outcomes, and are tied to workload in terms of training effectiveness. Three instructional strategies were analyzed in a between-subjects design for their degree of perceived workload: Highlighting, Massed Exposure, and Kim's game. Workload included subjective mental and global demand subscales from the NASA-Task Load Index (NASA-TLX). A multivariate analysis of variance showed that Kim's game contained the highest mental and global demands; Highlighting produced the lowest mental and global demands. The differences in demands suggest rationales for strategy placement, in terms of combinations (e.g., layers or progressions) and other applications (e.g., air traffic control, medical diagnosis, and After-Action Reviews). The strategies' workload differences are also traced to differences in attention and working memory.

**Keywords:** Training effectiveness · Instructional design for simulations · Workload · Signal detection · Virtual environments

## 1 Introduction

Soldiers can benefit from training their minds to identify inconspicuous events. In the military realm of human Intelligence, Surveillance, and Reconnaissance (ISR) tasks, the ability to gather valuable information on a high-value target is vital [1]. Today, Soldiers are required to observe human behavior in highly irregular, complex environments. Target cues include behavioral features that indicate the presence of a threat. A subset of target cues are kinesics, or non-verbal gestures that convey one's true emotional states [2]. To train one to adopt the standpoint of a Combat Hunter (where kinesics is an approach to gauge intent through suspicious behavior), three instructional strategies were applied to enhance Simulation-Based Training (SBT). Since the SBT application

is novel, workload between the strategies were analyzed to both further perceptual training and aid strategy selection for different objectives.

### **1.1 Threat Detection**

Threat detection requires an observer to allocate perceptual and attentional resources to decide if a target cue is present in the environment. Perceiving a threat involves detecting a signal (i.e., a target cue) amongst competing noise (i.e., distractions such as non-target cues) [3]. Thus, detection is enacted under some uncertainty, and involves discrimination of perceptual elements to make sense of an environment. The present study required users to detect a target cue, and then classify the type of target cue as signaling nervousness or aggressiveness. A person activates components of Working Memory (WM) when information is processed in the environment.

### **1.2 Working Memory**

WM is the mental ability to temporarily maintain and coordinate perceived stimuli necessary for performing higher-order cognitive operations. However, WM is limited in capacity: performance is impaired if one is required to process numerous external stimuli [4, 5]. Visual WM has been found to interact closely with attention during encoding and manipulating sensory stimuli. This has led researchers to postulate that attention is necessary for sustaining information [6]. Hence, when a Soldier processes incoming information from an external stimulus, he or she must undergo a multi-step psychological process, requiring both WM and attention. This process follows a loop of actions, including scanning for potential targets, target acquisition, analysis, decision-making, and disengagement. For the current experimental interest of a threat detection task for a simulated field of operations, a viewer compares instances of kinesic cues held in working memory to kinesic pattern knowledge solidified in long-term memory [7]. This process allows the user to determine if a given experimental cue matches a prototypical target cue. The role of working memory and attention can be adjusted by instructional strategies, and differing strategies are expected to vary in workload.

### **1.3 Workload**

Workload is the collective contributions of task processes, user resources, and environmental influences imposed on a person while completing a task [8]. The principle of workload has been extensively investigated across various domains, including medical, military, and civil aviation tasks. In most domains, multitasking is a crucial component and is impacted by workload. A red-line denotes where a person faces performance losses, and where unacceptable system characteristics exist [9, 10]. Effective amounts of workload can be viewed on a continuum, where workload levels should be neither too low nor too high. Extremes on this continuum would likely hinder learning; for example, motivation may be impacted adversely [11]. Further, empirical research has shown that high workload is associated with exhaustion [12], performance decrement [13], and task difficulty [14]. For strategy video games, Hsu, Wen, and Wu [15]

contended that the greatest challenge (resulting in fun, engaging gameplay) incorporates an intermediate amount of mental demand. In these lights, workload acts as a criterion for assessing users' reactions in terms of training's effectiveness. This underscores the need to examine and frame workload.

One measure that has achieved reliability and validity for capturing workload is the NASA-Task Load Index (NASA-TLX) [16]. The NASA-TLX is a multidimensional, self-report survey that assesses a user's subjective experience through six subscales [8]. These subscales include three demand subscales, as well as performance, effort, and frustration subscales. The three subscales of mental, physical, and temporal demand correspond to a user's cognitive expenditure, physical expenditure, and time-pressure strain, respectively. The performance subscale measures the user's perceived degree of task success, after a completed task. The effort subscale measures the user's feeling of difficulty due to resource input. The frustration subscale measures the degree of adverse emotions that a task triggers. Finally, global demand is an overall rating obtained by integrating all six domains, creating an average sum [17]. For this experiment, the NASA-TLX questionnaire was used as an investigative tool to assess perceptions of mental demand and global demand within a Virtual Environment (VE) for perceptual training of kinesic cues.

#### **1.4 Virtual Environments**

VEs are synthetic representations of a reality, where the user perceives that he or she is operating within the synthetic reality [18]. VEs are an economical alternative to training exclusively in real environments (e.g., practicing detection of improvised explosive devices), and benefit from the customization of complex virtual scenarios that are not readily available to a user (e.g. analysis at an atomic level). VEs may also allow for an After-Action Review (AAR) to aid in skill expansion [19, 20]. Ultimately, the benefits of VEs afford ways to improve instruction.

A VE is a platform for accommodating Simulation-Based Training (SBT). SBT is an approach to training that enables a user to focus mental and/or physical resources in order to accomplish a task under differing environmental conditions (e.g. a pilot using a computer for various flight training scenarios). A training simulation offers replication of a task's corresponding cognitive demands, thereby furthering skill acquisition through virtual experience of the task [21]. An enhanced pedagogical approach to SBT is through the implementation of instructional strategies. For detection tasks, specific focus is given to perceptual instructional strategies.

#### **1.5 Perceptual Instructional Strategies**

Perceptual instructional strategies are designed to improve a user's observation skills [22]. Based on perceptual, signal-detection needs, three valid strategies for kinesic cue detection were identified for the present analysis: Highlighting, Massed Exposure, and Kim's game.

Highlighting is the purposeful use of a distinct stimulus to direct a user's visual resources toward a feature, characteristic, or object of interest within a simulated

environment. In order to orient a user to a point of interest, a stimulus must contrast (e.g., through the use of color) from the surrounding environment. The stimulus' distinguishing characteristic allows a user to recognize a highlighted feature [23].

Massed Exposure (ME) depicts multiple signals, or target cues, for a user to process simultaneously. A raised signal-to-noise ratio is sought to saturate the trainee with opportunities for skill practice. The increased presence of both target cues and noise suggests the need for selective attention in order to mediate between the different cues.

Kim's game is designed to heighten several types of cognitive skills, including enhanced sensitivity in change detection [24]. Like threat detection, change detection is a perceptual phenomenon crucial to ISR: soldiers must note changes in an operational environment, and then make sense of such changes. Kim's game stems from the novel *Kim*, where the main character underwent memorization training in a two-step process: initial observation of a set of objects for a period of time, and subsequent recollection of the same objects from memory [25]. The game mechanic consisted of how well Kim could remember the objects' features after they were veiled from sight. A variation of the game, which involves change detection, presents another set of items after the veiling period (i.e., an Interstimulus Interval (ISI)) and requires the player to detect if an item changed from the original set. This latter variation was virtually presented in the experiment.

## 1.6 Purpose

Threat detection and change detection require several cognitive mechanisms that contribute to workload. A Soldier conducting detection tasks must strive to allocate mental resources optimally. The application of instructional strategies within a VE offers enhanced training features that cannot be replicated easily in a real-world environment. The purpose of this research initiative is to analyze how each instructional strategy contributes to user workload. Given workload's relation to training effectiveness, one research question is to determine the placement and application of each strategy. The elucidation of a task-to-strategy goodness of fit through workload analysis also has potential beyond threat detection training, based on contextual performance needs.

## 2 Hypotheses

The first hypothesis links mental demand and utilization of a visual highlighting cue. A non-content, extradiegetic visual cue can act as a tool to assist a user in quickly locating a specific target. The quick orientation afforded by the cue suggests a low degree of mental demand, due to a reduction in a user's vigor to scan a scene. Jerome, Witmer, and Mouloua [26] discovered that visual cues contributed significantly to improved detection accuracy when a participant was under a high amount of workload during a search task. In our treatment, Highlighting reduced the tentative aspect of target detection by its non-content feature, leaving only the subtask of cue classification (i.e., nervousness or aggressiveness) to be performed. Accordingly, these links lead to hypothesis one.

**Hypothesis 1 (H1):** Highlighting produces the lowest mental demand and lowest global demand compared to ME and Kim's game.

The second hypothesis is based on memory recall. A discrete change detection task involves an interruption between visual scenes through an ISI (or break); a user is forced to recall a previous scene's information, and compare the information to a newly presented scene. Change detection not only focuses on processing an observed change, but detecting where the change occurred [27]. The process for solving the observational puzzle of Kim's game includes two strategy options: comparing scene-to-scene features to detect an anomaly, and comparing a given scene's kinesis cues to kinesis knowledge. Only the latter is available for Highlighting and Massed exposure. In contrast, the cooperation of a double-strategy from Kim's game may lower mental demand in comparison to Massed Exposure. Yet, unlike Highlighting, an unaided detection task still exists to increase mental demand.

A higher degree of temporal demand, effort, and frustration is presumed to exist from the discrete change detection task. The user is not given time to immerse themselves in the microworld, due to the ISI's visual discontinuity, and thus he or she is more aware of time pressure. The unnatural presentation of cues may induce effort and/or frustration, since the user must disengage with the virtual environment at the onset of the ISI, and re-engage with the virtual environment after the ISI. These considerations lead to hypothesis two.

**Hypothesis 2 (H2):** Kim's game produces higher mental demand than Highlighting, but less mental demand than ME; and produces higher global demand than both Highlighting and ME.

Both Kim's game and Massed Exposure have an identical target cue ratio (2:3), and a similar number of total cues presented in the scenarios. Thus, targets alone are not expected to differ in inducing mental demand, and other characteristics of each instructional strategy must be considered. In terms of perceptual load for detection, each model in the ME condition must be examined by the user to accept or reject a cue match. However, Kim's game does not have this limit: a user compares scenes, but does not necessarily have to look at every single cue to make a decision. Viewing discrete scenes in separate trials for Kim's game may also demand less vigilance (and mental demand) than ME's perpetual, continuous task. According to the current expectations, Kim's game has the highest global demand, and Highlighting the lowest; this order leaves ME with a medium-level of global demand. Based on these ideas, hypothesis three is formed.

**Hypothesis 3 (H3):** ME produces higher mental demand than both Highlighting and Kim's game; but produces only less global demand than Kim's game.

## 3 Method

### 3.1 Participants

Individuals were recruited through flyers and emails distributed to the University of Central Florida community. Participants were required to meet specific inclusion criteria in order to participate, which included a United States citizenship, a minimum of 18 years of age, normal (or corrected-to-normal) vision, and a lack of participation in

previous SBT experiments. Each participant received a choice for compensation: either monetary compensation or class credit for up to five hours.

In terms of participant demographics, the Highlighting group consisted of  $n = 37$  with an age range of 18–32 ( $M = 22.68$ ,  $SD = 3.53$ ), the ME group consisted of  $n = 32$  with an age range of 18–33 ( $M = 22.22$ ,  $SD = 2.81$ ), and the Kim's game group consisted of  $n = 34$  with an age range of 18–38 ( $M = 22.18$ ,  $SD = 4.41$ ).

### 3.2 Experimental Design

A between-subjects design was utilized with one independent variable: the instructional strategy. The independent variable consisted of three conditions: ME, Highlighting, and Kim's game. The two dependent variables were the NASA-TLX mental demand and global demand scores.

For each training strategy, users were to detect and classify aggressive and nervous cues performed by animated models. Non-target cues were intermixed with the target cues. Aggressive cues comprised models that either clenched their fists, or slapped their hands. Nervous cues comprised models that either wrung their hands, or looked behind their back to "check six o'clock." Non-target cues consisted of idle talking, crossed arms, a rubbed neck, or the check of one's watch. Per every three events (i.e., instance of a cue) shown in the Highlighting condition, one event would represent a target cue. Further, this target cue was overlaid with a transparent blue box. Per every three events shown in the ME condition, two would represent a target cue. For both of these conditions, the kinesic models were viewed during the simulation of an unmanned vehicle's traversal through a town. That is, the user took on the role of an operator of a continuously moving robotic vehicle, where models were populated along the vehicle's route.

Kim's game required a user to first observe a baseline scene of non-target models for eight seconds. Second, the user saw a 1-second long ISI, as a black screen. Third, the user was presented with another scene (for eight seconds), and was required to use working memory for change detection: the user was required to detect if a non-target cue model from the first scene changed to a target cue model in the second scene (within a 20 s time range). Between scenes, the models may or may not have changed; if a change occurred, it would be present in only one model. Kim's game incorporated the 2:3 target ratio chance that a target model would appear. However, this condition incorporated an immobile camera's point-of-view of a town, instead of simulating a continuously moving vehicle.

### 3.3 Surveys

The demographics questionnaire gathered basic descriptive information about each participant. This information included age, gender, education, and experience with computers.

The NASA-TLX (Hart and Staveland, 1988) served as the measure for workload. The NASA-TLX contained a scale ranging from 0 to 100 for each subscale. A rating of 0 indicated a "very low" rating for the corresponding subscale. A rating of 100 indicated a "very high" rating for the corresponding subscale. An exception applied to the performance subscale, as a 0 indicated a "perfect" rating and a 100 indicated a "failure" rating.

### 3.4 Materials

A 22-inch desktop computer with a 16:10 aspect ratio was used across all three conditions. The Virtual Battlespace 2 (VBS2) software program was selected for developing and administering the experimental scenarios.

### 3.5 Procedure

The participant, upon arrival to a predetermined lab location, was greeted by the experimenter. The experimenter cross-checked to ensure the participant had not been involved in previous SBT studies. If the participant was allowed to continue, he or she was randomly assigned to one of the three conditions. For all three conditions, environmental workload was controlled for by allocating lab space for each participant. After the participant was assigned to a condition, they completed an informed consent form. The informed consent was followed by the Ishihara Test for Colour Blindness [28] to ensure continued participation. Next, the demographics questionnaire was administered.

Following the demographics questionnaire, interface training was provided to each participant in order to familiarize them with system navigation and detection methods. Each participant then experienced a pre-test scenario. The scenario was intended to test a participant's prior knowledge of nervous and aggressive kinesic cues. A second iteration of interface training followed for all conditions. In general, the participant was asked to detect and classify yellow- and red-colored target barrels, distinct from non-target barrels of different colors. The Highlighting condition presented translucent blue boxes over the red and yellow barrels. The ME condition presented twice as many target barrels as the Highlighting condition, but lacked the non-content highlighting feature. The Kim's game preparation operated similarly to its corresponding kinesic-training form, with colored barrels instead of cue models. In all conditions, the implementation of barrels circumvented priming for the instructional material of kinesic cues.

Following the second interface training, the experimenter presented a PowerPoint demonstrating target behavior cues. The participant then completed the instructional training task, with the objective to identify target cues in the VE for the given condition. The NASA-TLX questionnaire was administered after the training scenarios. A 5-minute break followed the completed questionnaire. Afterward, the experimenter presented a slide deck as interface training for the post-test, and then administered the post-test scenario. Finally, the participant was thanked, debriefed, and dismissed. The experiment lasted up to 5 h. The research study was conducted through a multi-year effort.

## 4 Results

### 4.1 Assumption Testing

A between-groups one-way multivariate analysis of variance (MANOVA) was performed. The statistical assumptions of the MANOVA were investigated to ensure validity in the results. Testing with box-and-whisker plots, only one value was discovered as a potential outlier for the NASA-TLX mental demand scores. However, it was

retained, since the  $z$  score was determined to be within 3 standard deviations from the mean. A Lilliefors test of normality revealed non-significance, indicating distributions were approximately normal. A regression analysis revealed no violations of Mahalanobis distances exceeding critical values. A Levene's test confirmed no violations across each of the groups, which supported the presence of homogeneous variance. Finally, a Box test indicated equality of the covariance matrix.

## 4.2 MANOVA Results

Results of the MANOVA revealed significance between the instructional strategies and the dependent variables of mental and global demand  $F(6, 262) = 10.08, p < .001$ ; a Pillai's trace = .375; partial eta squared = .19. Pillai's trace was selected due to its resistance against any undesired effects that have been introduced by the variations in sample sizes among the independent variable [29]. A Bonferroni adjustment for the significance level was implemented to determine additional significance by each dependent variable. The adjusted significance level of .025 revealed significance for mental demand  $F(3, 131) = 18.866, p < .001$ , partial eta squared = .30 and global demand  $F(3, 131) = 21.91, p < .001$ , partial eta squared = .33.

A post hoc analysis (Tukey HSD) was conducted through two separate one-way analyses of variance (ANOVAs). An analysis of mental demand indicated statistically significant differences existed between the Highlighting, ME, and Kim's game groups. An analysis of global demand indicated significant differences between the Highlighting, ME, and Kim's game groups. An additional analysis of the means indicated that Kim's game contained the highest mean for mental demand ( $M = 69.56, SD = 19.04$ ) and global demand ( $M = 45.00, SD = 11.98$ ).

## 5 Discussion

**H1:** Highlighting produces the lowest mental demand and lowest global demand compared to ME and Kim's game.

Hypothesis 1 was fully supported. The use of Highlighting for behavior cue detection could serve as a multipurpose tool. Highlighting's lower degree of mental demand suggests that it may serve to compliment other strategies that require a higher degree of mental resources (e.g. Highlighting may be layered with Kim's game). The application of Highlighting in high-resource strategies could serve to improve a Soldier's cognitive skillsets in change detection training. As an example, a Highlighting strategy could require a novice user to methodically scan a visual scene, as is common in sniper fieldcraft and improvised explosive device detection. The user would first practice following an expert-level visual path with an identifiable cue. Then, the user would later perform the scan, without the cue, in the same methodical fashion.

The results add empirical support for the analysis by Carroll, Milham, and Champney [22], who suggested that Highlighting may be used effectively to assist in an AAR. AARs often serve as a way for a user or trainee to receive feedback from a more skilled instructor. The low degree of global demand for Highlighting suggests it is an ideal method for

enhancing AAR feedback. A user could easily orient toward a specific area of visual interest, while an instructor coaches the user.

**H2:** Kim's game produces higher mental demand than Highlighting, but not less mental demand than ME; and produces higher global demand than both Highlighting and ME.

Hypothesis 2 received partial support. As expected, global demand was highest for Kim's game, likely due to the requirements of the strategy (e.g. temporal demand and effort). The high degree of mental demand for Kim's game suggests that the number of targets or stimuli a user must process simultaneously may not necessarily be the primary influence on his or her mental demand rating. The concept of having two strategies may have had the reverse effect predicted, and created additive workload; or perhaps only one strategy was used, and this one strategy was more complex than that used by the Massed Exposure group participants. That is, the Massed Exposure group's strategy for detection was more automatic. A user's perception of mental demand may be influenced predominately by the WM demands of the instructional strategies, since a user is consciously aware of WM processes (for a review, see Estes [30]). Learned perceptual processes are more automatic, which reduces the likelihood that the user is aware they are engaged in information processing. The conscious awareness of items held in WM could account for the high degree of mental demand found in Kim's game. Through consideration of the amount of mental demand a task requires, instructors may inform their task-to-strategy requirement decisions: if a Soldier is a novice to threat detection training, transfer of learning from SBT to the real-world could be facilitated by introduction with a low-workload strategy. As learner proficiency increases, the instruction may progress to a higher-level workload strategy.

**H3:** ME produces higher mental demand than both Highlighting and Kim's game, but produces only less global demand than Kim's game.

Hypothesis 3 received partial support. Contrary to the original hypothesis, mental demand was not highest for ME. It appears that ME relies on selective attention, which is dependent on the amount and type of load imposed by a task [31]. A user requires working memory as a primary source of mental demand in Kim's game. Target detection using ME may have become an automatic process, which reduced cognitive resources beyond that of Kim's game. Additionally, this process may be a consequence of the user's pre-exposure to the perceptual stimuli during the pre-test scenario. Although each strategy began with a pre-test scenario, ME is highly similar to the pre-test scenario than that of Kim's game. Research by Basile and Hampton [32] suggests that recognition of a task is much easier and faster than recall. Thus, familiarity of the task stimuli (due to pre-exposure in the pre-test scenario) may have reduced challenges and difficulties related to mental demand.

Unsurprisingly, ME's effect on participant's global demand was lower than that of Kim's game, and higher than that of Highlighting. This may be explained by the increased effort, frustration, and/or temporal demand required for the Kim's game task. Perhaps, the ME strategy required less effort because the task was less exhausting and the scenario itself was continuous and appeared to be seamless. The temporal separation between interdependent scenes could cause confusion. These explanations help to support the reduced

overall global workload participant scores in the ME group, in comparison to Kim's game participants.

Adopting ME as an instructional design application for detection training can be beneficial in areas where an individual's visual load is saturated with stimuli. ME has the potential to increase focus for target identification amongst distraction. For example, an air traffic controller may benefit from training using ME to improve detection accuracy for identifying an airplane in imminent danger. In addition, the medical field could benefit from using ME training, because it would train professionals to be more accurate in detecting abnormalities in X-rays and MRI scans. Further research on brain activity (and other physiological measures) during threat detection could provide additional findings to explain and compare the demands of the strategies, further defining their role in future training.

### 5.1 Additional Considerations

Previous studies, which assessed performance aspects of the present experiment, add depth to understanding workload differences. Notably, post-test detection accuracy scores did not significantly differ between Highlighting and Massed Exposure conditions; nor did detection accuracy percent-change scores (i.e., change in detection accuracy from pre-test to post-test) [33]. Similarly, post-test detection accuracy scores did not significantly differ between Kim's game and Massed Exposure conditions; nor did detection accuracy percent-change scores [7]. In other words, none of the strategies compared had a substantial impact on detection performance. This lack of objective difference leads to considering measures of subjective workload as insightful criteria. According to our results, Highlighting has a more efficient use of workload than ME, and ME has a more efficient use of workload than Kim's game (no study has compared performance differences between Highlighting and Kim's game).

Although the NASA-TLX provides a reliable measure of general workload, the survey is insensitive to denote types of load formalized in the Cognitive Load Theory (CLT). Within CLT, load drivers are divided into three kinds, all of which relate to working memory demands: intrinsic load deals with the nature of information to be learned, and the previous domain knowledge of a user; extraneous load deals with aspects of instructional design irrelevant to knowledge acquisition; and germane load deals with the amount of resources a user directs toward learning [34]. As such, future investigation should consider whether load was incurred in an intrinsic, extraneous, or germane way. For example, Kim's game may have had high workload due to extraneous load (e.g., the temporal separation of model references may have induced high demands on working memory) or germane load (e.g., the puzzle was highly engaging, and this frustration-as-fun effect drew motivational resources for learning).

## 6 Conclusion

As an initial empirical assessment of cue detection training through a virtual simulation, this study framed workload demands from lowest to highest, in terms of three strategies'

mental and global demands. Differences of workload were determined at significant levels for Highlighting, Massed Exposure, and Kim's game strategies. Their distinctions were contextualized by functions of human perception, such as attention and working memory. Further, the differences in workload afford a designer a systematic selection rationale for learning activities. For applications, Highlighting suggests a way to scaffold (e.g., scan-path training) and a way to reinforce feedback; Kim's game may be suitable for high-challenge needs; and Massed Exposure lends itself to professionals (e.g., air traffic controllers, doctors analyzing X-rays) for practice of crucial decisions via saturated event instances.

**Acknowledgments.** This research was sponsored by Dr. Irwin Hudson of the U.S. Army Research Laboratory Human Research Engineering Directorate Advanced Training and Simulation Division (ARL HRED ATSD), under contract W911NF-14-2-0021. However, the views, findings, and conclusions contained in this presentation are solely those of the author and should not be interpreted as representing the official policies, either expressed or implied, of ARL HRED ATSD or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government.

## References

1. Fautua, D., Schatz, S., Kobus, D., Spiker, V.A., Ross, W., Johnston, J.H., Reitz, E.A.: Border Hunter Research Technical Report. U.S. Joint Forces Command, Norfolk (2010)
2. Birdwhistell, R.L.: *Kinesics and Context: Essays on Body Motion Communication*. University of Pennsylvania, Philadelphia (1970)
3. Wickens, C.D.: *Engineering Psychology and Human Performance*. HarperCollins, New York (1992)
4. Baddeley, A.D., Hitch, G.: Working memory. In: Bower, G.H. (ed.) *The Psychology of Learning and Motivation: Advances in Research and Theory*, vol. 8, pp. 47–89. Academic Press, New York (1974)
5. Awh, E., Barton, B., Vogel, E.K.: Visual working memory represents a fixed number of items regardless of complexity. *Psychol. Sci.* **18**(7), 622–628 (2007)
6. Fougne, D.: The relationship between attention and working memory. In: Johansen, J.B. (ed.) *New Research on Short-Term Memory*, pp. 1–45. Nova Biomedical Books, New York (2008)
7. Maraj, C.: *Investigating simulation-based pattern recognition training for behavior cue detection* (Doctoral dissertation) (2015). <http://stars.library.ucf.edu/etd/1285/>. Accessed
8. 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., Hancock, P.A., Meshkati, N. (eds.), *Human Mental Demand*, pp. 139–183. Oxford, North-Holland (1988). doi:[10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
9. Colle, H.A., Reid, G.B.: Estimating a mental workload redline in a simulated air-to-ground combat mission. *Int. J. Aviat. Psychol.* **15**, 303–319 (2005)
10. Grier, R., Wickens, C., Kaber, D., Strayer, D., Boehm-Davis, D., Trafletton, J.G., John, M.S.: The red-line of workload: theory, research, and design. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **52**(18), 1204–1208 (2008). Sage Publications
11. Kyndt, E., Dochy, F., Struyven, K., Cascallar, E.: The direct and indirect effect of motivation for learning on students' approaches to learning through the perceptions of workload and task complexity. *High. Educ. Res. Dev.* **30**(2), 135–150 (2011)

12. Bentzen, M., Lemyre, P.N., Kenttä, G.: Development of exhaustion for high-performance coaches in association with workload and motivation: a person-centered approach. *Psychol. Sport Exerc.* **22**, 10–19 (2016)
13. Yurko, Y.Y., Scerbo, M.W., Prabhu, A.S., Acker, C.E., Stefanidis, D.: Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. *J. Soc. Simul. Healthc.* **5**(5), 267–271 (2010)
14. Brookings, J.B., Wilson, G.F., Swain, C.R.: Psychophysiological responses to changes in workload during simulated air traffic control. *Biol. Psychol.* **42**(3), 361–377 (1996)
15. Hsu, S.H., Wen, M.H., Wu, M.C.: Exploring design features for enhancing players' challenge in strategy games. *CyberPsychology Behav.* **10**(3), 393–397 (2007)
16. Hart, S.: NASA-Task load index (NASA-TLX): 20 years later. *Proc. Hum. Factors Ergon. Soc. Annual Meet.* **50**(9), 904–908 (2006). Sage Publications, Los Angeles
17. Grier, R.A.: How high is high? a meta-analysis of NASA-TLX global demand scores. *Proc. Hum. Factors Ergon. Soc. Ann. Meet.* **59**(1), 1727–1731 (2015). Sage Publications, Los Angeles
18. Sheridan, T.B.: Defining our terms. *Presence Teleoperators Virtual Environ.* **1**(2), 272–274 (1992)
19. Todorov, E., Shadmehr, R., Bizzi, E.: Augmented feedback presented in a virtual environment accelerates learning of a difficult motor task. *J. Mot. Behav.* **29**(2), 147–158 (1997)
20. Eschenbrenner, B., Nah, F.F., Siau, K.: 3-D virtual worlds in education: applications, benefits, issues, and opportunities. *J. Database Manag. (JDM)* **19**(4), 91–110 (2008)
21. Cannon-Bowers, J.A., Bowers, C.: Synthetic learning environments: on developing a science of simulation, games, and virtual worlds for training. *Learn. Training Dev. Organ.*, 229–261 (2009)
22. Carroll, M., Milham, L., Champney, R.: Military observations: perceptual skills training strategies. In: *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, Arlington, VA: NTSA (2009)
23. Posner, M.I.: Orienting of attention. *Q. J. Exp. Psychol.* **32**(1), 3–25 (1980). doi:[10.1080/0033558008248231](https://doi.org/10.1080/0033558008248231)
24. Maraj, C.S., Lackey, S.J., Badillo-Urquiola, K.A., Hudson, I.L.: Assessment of Kim's game strategy for behavior cue detection: engagement, flow, & performance aspects. In: Lackey, S., Shumaker, R. (eds.) *Virtual, Augmented and Mixed Reality. VAMR 2016. Lecture Notes in Computer Science*, vol. 9740, pp. 156–163. Springer, Cham (2016). doi:[10.1007/978-3-319-39907-2\\_15](https://doi.org/10.1007/978-3-319-39907-2_15)
25. Kipling, R.: *KIM*. Macmillan & Co., London (1901)
26. Jerome, C., Witmer, B., Mouloua, M.: Attention orienting in augmented reality environments: effects of multimodal cues. *Proc. Hum. Factors Ergon. Soc. Ann. Meet.* **50**(17), 2114–2118 (2006). Sage Publications
27. Rensink, R.A.: Change detection. *Annu. Rev. Psychol.* **53**(1), 245–277 (2002)
28. Ishihara, S.: *Ishihara's Tests for Colour Deficiency*. Kanehara Trading, Tokyo (2013)
29. Olson, C.L.: On choosing a test statistic in multivariate analysis of variance. *Psychol. Bull.* **83**, 579–586 (1976). doi:[10.1037/0033-2909.83.4.579](https://doi.org/10.1037/0033-2909.83.4.579)
30. Estes, S.: The workload curve: subjective mental demand. *Hum. Factors* **57**(7), 1174–1187 (2015). doi:[10.1177/0018720815592752](https://doi.org/10.1177/0018720815592752)
31. Burnham, B.R., Sabia, M., Langan, C.: Components of working memory and visual selective attention. *J. Exp. Psychol. Hum. Percept. Perform.* **40**(1), 391–403 (2014). doi:[10.1037/a0033753](https://doi.org/10.1037/a0033753)
32. Basile, B., Hampton, R.: Recognition errors suggest fast familiarity and slow recollection in rhesus monkeys. *J. Learn. Mem.* **20**(8), 431–437 (2013). doi:[10.1101/lm.029223.112](https://doi.org/10.1101/lm.029223.112)

33. Salcedo, J.N.: Instructional strategies for Scenario-Based Training of human behavior cue detection and classification with Robot-Aided Intelligence, Surveillance, and Reconnaissance. (Doctoral dissertation). (2014). <http://stars.library.ucf.edu/etd/1304/>. Accessed
34. Sweller, J., Ayres, P., Kalyuga, S.: Cognitive Load Theory. Springer, New York (2011). doi: [10.1007/978-1-4419-8126-4](https://doi.org/10.1007/978-1-4419-8126-4)