

# The Implementation of Instructional Strategies for Training in a Virtual Environment: An Exploratory Investigation of Workload and Performance

Elizabeth Wolfe, John Granger, Natalie Alessi, Lauren Farrell,  
and Crystal Maraj<sup>(✉)</sup>

Institute for Simulation and Training,  
University of Central Florida, Orlando, FL, USA  
{ewolfe, jgranger, nalessi, lfarrel, cmaraj}@ist.ucf.edu

**Abstract.** This study expanded upon the use of Simulation-Based Training (SBT) to enhance kinesic cue detection in military operations. The research investigated three SBT instructional strategies: Massed Exposure (ME), Highlighting, and Kim's Game. An additional Control condition was included as a validity measure for the tested instructional strategies. Each strategy provided a separate perceptual experience for learning how to detect the target kinesic cues specified: aggressiveness and nervousness. The subjects' workload scores from the NASA-Task Load Index (NASA-TLX) subscales of performance, effort, and frustration provided greater understanding of each strategy's impact on perceptual skill improvement. For each instructional strategy, this research increased understanding of workload efficiencies and highlighted potential applications. Examination of descriptive statistics revealed that Kim's Game scored the lowest in detection accuracy, and exceeded both Highlighting and ME in terms of effort, performance, and frustration on the NASA-TLX. Highlighting scored highest in detection accuracy, but lowest in effort, performance, and frustration. The Control and ME scored neither highest nor lowest for all NASA-TLX subscales and detection accuracy. Given the results, the instructional strategies may be implemented individually or used in combinations for a desired training effect. Kim's Game may be suited for training experts, whereas Highlighting and ME may be suited for novice trainees. Additional fields, such as baggage screening, healthcare, and education may benefit from implementation of workload-appropriate perceptual training strategies within a virtual environment.

**Keywords:** Simulation-Based Training (SBT) · NASA-TLX · Workload · Virtual environment · Kinesic cues

## 1 Research Foundation

A Warfighter must accurately detect and interpret threats in the field. One type of threat is in the form of kinesic cues, here nonverbal body movements communicating one's intent [1]. The target kinesic behavior cues conveyed in this study were either

aggressive or a nervous behaviors signifying a potential threat. Aggressive behaviors included clenching one's fists or slapping one's hands, while nervous behaviors included checking one's "six o'clock" or wringing one's hands (i.e., clasping and rubbing the palm of one hand with the other hand). Currently, a need exists to increase accuracy in detecting threats in a Warfighter context. This need has inspired the use of instructional strategies in a virtual environment for training perceptual skills.

## 1.1 Instructional Strategies

Instructional strategies were sought to aid knowledge acquisition of target cues in a virtual environment. The instructional strategies included Kim's Game, Highlighting, and Massed Exposure (ME). Kim's Game was adopted from Rudyard Kipling's book titled *Kim*, where the former focused on teaching change detection and recall. Adopting the premise behind Kim's Game allowed for the creation of a virtual version presented in the form of an instructional strategy. The strategy presented a discrete task whereby the trainee had to identify a change (or no-change) in a virtual scene. Specifically, the trainee observed a scene of non-target cues for 8 s, a blank screen for 1 s (i.e., an interstimulus interval), and a second scene of cues for 8 s. The second scene required change detection, as one cue may have changed from a non-target to a target cue. Following, the trainee chose whether a change had occurred or not within final scene. This tasks the trainee with recalling the first stimuli, and then detecting whether a change has occurred [2].

Another instructional strategy used in this study was Highlighting, which is the use of a non-content stimuli within the simulated environment to orient the trainee's attention toward the target cue. A translucent blue box appeared over the target cues, leaving the user to classify the presence of an aggressive or nervous cue. ME involves the saturation of practice opportunities by presenting a higher amount of the target cues to the user, in comparison to the Control group. A Control condition was incorporated as a continuous task with no formal training (i.e., the absence of an instructional strategy). Overall, each strategy requires the selective attention of the trainee, in order for the trainee to accurately detect the target cue among non-target cues [3].

## 1.2 NASA-Task Load Index

This research utilized the NASA-Task Load Index (NASA-TLX) survey, which is a subjective tool used to assess perceived workload for a given task. The full survey consists of six subscales, including mental demand, physical demand, temporal demand, performance, effort and frustration [4]. Each subscale is comprised of 100 points, with a 5-point interval step. For this report, the survey comprises effort, frustration, and performance subscales. Specifically, effort referred to how much physical and mental strength was needed to accomplish the task. Frustration examined the level of stress, versus relaxation, felt during the task. Finally, performance described the level of confidence the participant had in their performance of the task. Unlike the other subscales, performance was scored as '0' to represent perfect performance and '100' to

represent lowest performance. The mental demand subscale, as well as a global demand scale, were evaluated and presented in research by [5] and will not be duplicated in this research effort.

### 1.3 Detection Accuracy

For selecting correct target cues, detection accuracy was logged and scored. The score was measured by the percentage of correctly detected targets cues out of the total number of target cues presented.

### 1.4 Research Questions

This research explores the impact of workload responses (i.e., effort, frustration, and performance) using a between-subjects design for the instructional strategies (i.e., Kim's Game, ME, Highlighting and Control). Changes in cognitive demands and resource allocation may result in differences reported for workload, and ultimately detection accuracy. This has inspired an investigation into the following research questions:

1. Which strategy will score highest or lowest for effort, frustration, performance, and detection accuracy?
2. How do the strategies differ in workload and detection accuracy?
3. Which strategy is most effective in yielding an appropriate amount of workload?

## 2 Data Collection

The experimenter greeted each participant and verified that he or she met the inclusion requirements (i.e., the participant was a U.S. citizen, was at least 18 years of age, had normal or corrected vision, and had not participated in prior Simulation-Based Training experiments). After entering the experiment room, the experimenter obtained the participant's signed consent and administered the color blindness test [6]. If the participant passed the color blindness test, he or she was assigned to one of the four conditions.

Next, the experimenter asked the participant to view a scenario (on a 22-inch desktop computer) which familiarized him or her with the user interface and the testbed's Virtual Battlespace 2 (VBS2) software. Following this event, the experimenter requested that the participant complete a pre-test scenario, using prior experience to detect and classify cues as exhibiting either aggressiveness or nervousness. After, the experimenter presented a PowerPoint training on behavior cue detection with explanations for aggressive and nervous cues. The presentation also highlighted non-target cues (such as idle talking and cross arms) as well as instructions for completing the task. Next, the experimenter displayed one of the four conditions (i.e., Kim's Game, ME, Highlighting or Control) on the desktop and asked the participant to complete the scenario. Following the scenario, which was displayed for approximately

15 min, the experimenter requested that the participant complete the on-line version of the NASA-TLX.

After the NASA-TLX was completed, a final PowerPoint with instructions to complete the post-test was presented by the experimenter. Afterwards, the participant completed the post-test, which tested knowledge acquisition and logged the detection accuracy scores used in the present analysis. Following this, the participants were debriefed, compensated (with either \$10 per hour or class credit), and dismissed. The total duration of the experiment was roughly 3 h.

### 3 Results

Preliminary analyses indicated that the age ranges of the participants were 18–38 ( $M = 22.12$ ,  $SD = 3.44$ ). The number of both male and female participants was 65, with a sample size of 130. Closer inspection of the data showed that the distribution was non-normal, or nonparametric, in nature. Therefore, it was more meaningful to examine the medians for the NASA-TLX subscales (i.e. effort, frustration and performance) and detection accuracy. Observing the median distribution for all the conditions, Kim's Game had the highest median scores for effort (60), performance (35), and frustration (45) subscales, and had the lowest detection accuracy score (85.94%). Highlighting had the lowest median scores for the three workload subscales of effort (14), performance (3), and frustration (5), and had the highest score for detection accuracy (96.71%). ME scored neither the highest nor lowest in effort (47), performance (24), frustration (30) or detection accuracy (94.19%). Finally, the Control was neither highest nor lowest in effort (53), performance (17), frustration (26), or detection accuracy (94.57%).

As previously mentioned, the non-normal distribution prompted an inspection of the data to identify extreme outliers. A closer look at the detection accuracy scores revealed extreme outliers in three out of the four groups: select ME, Highlighting, and Control scores were beyond the lower outer fence. Additionally, Kim's Game's effort subscale showed extreme outliers beyond the lower outer fence. Finally, Highlighting's performance subscale displayed extreme outliers beyond the upper outer fence.

### 4 Discussion

Based on the results, responses to the following research questions are provided.

#### 4.1 Which Strategy Will Score Highest or Lowest for Effort, Frustration, Performance and Detection Accuracy?

Kim's Game scored highest in effort, frustration, and performance; but lowest in detection accuracy. The discrete nature of the Kim's Game target cue presentation may have stimulated more attention, which increased effort and frustration levels. As a result, performance scores were impacted because the user felt challenged by Kim's Game. When examining detection accuracy, it appears that the Control had a higher

positive impact than Kim's Game. In addition, the Control was comparable to ME in terms of detection accuracy. Overall, from a design perspective, a baseline helps determine if a strategy is warranted for downstream training acquisition.

Highlighting scored lowest in effort, frustration, and performance; but highest in detection accuracy. The subscale scores suggest that Highlighting was perceived as a simpler task than Kim's Game. Prompting towards the target cue in Highlighting may explain this perception, leading to lower scores in effort and performance. Thus, Highlighting may be beneficial for training users who require low-frustration learning environments. For example, children with autism may benefit from training that starts with the least intrusive approach [7].

#### **4.2 How Do the Strategies Differ in Workload and Detection Accuracy?**

Each set of workload and detection accuracy scores for the three instructional strategies differ, signifying distinct trends to inform future strategy selection. At first glance, Highlighting seems equivalent to the Control in terms of detection accuracy. However, there is a distinction between performance scores for the Highlighting and Control conditions (see results for an overview). The differences may be explained by increased task confidence, induced by the non-content stimuli. Highlighting, therefore, may be applied toward inexperienced trainees, to increase their confidence levels. Highlighting may be applied to first-year nurses to gain confidence in technical skills required for treating patients in stressful environments [8].

When ME is compared to Highlighting, considerable differences in workload exist for all subscale scores. Yet, differences in detection accuracy scores between the two strategies were minimal. Given the differences in workload subscales, but similarities in detection accuracy scores, ME can assist with training tasks that are neither too challenging nor too simplistic. ME may be suited for high-stake security fields that require a tax on effort, but need a high detection accuracy. To illustrate, ME may be introduced to baggage-screening training, which requires the screener to identify Improvised Explosive Devices (IEDs) [9].

#### **4.3 Which Strategy is Most Effective in Yielding an Appropriate Amount of Workload?**

The appropriate amount of workload for an instructional training strategy depends on the desired application. Tasks that entail a large amount of workload for adequate performance may use Kim's Game. Whereas training instances that demand execution of simple tasks may gear toward Highlighting. Alternatively, ME may be used as an intermediate training strategy that provides extensive amounts of practice, with moderate workload demands. The placement of instructional strategies within a training continuum can cater to specific trainee needs. For example, Highlighting may act as a simplified approach to train detection. Once the trainee has mastered basic task components, ME may be presented for practice purposes, to create repetition and moderate workload demands. Lastly, Kim's Game may act as a testing mechanism for detection

skills. Kim's Game is an appropriate testing agent because it generates high workload demands expected in the real world. Therefore, the presence of high detection accuracy in Kim's Game suggests ideal performance under high-workload conditions.

## 5 Limitations

A limitation of this study was the discrete aspect of Kim's game, when compared to the continuous nature of Highlighting, ME, and Control. This ecological difference in design may have acted as an extraneous variable, given that the pre-test and post-test also used a continuous method. Furthermore, outliers were detected in all conditions, which may have skewed the data.

## 6 Conclusion

This study analyzed the descriptive statistics of three instructional strategies, for the NASA-TLX subscales of frustration, effort, and performance; and detection accuracy. Based on the research findings, recommendations were given to map strategies according to a training continuum. Potential applications include baggage screening, special education, and healthcare. Future directions of research includes a follow-up study to assess learning retention for detection skills, and assessing other subscales of the NASA-TLX. Finally, to improve the objectivity of the results, physiological measurements to assess trainee workload should be integrated.

**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. Special thanks to Mr. Jonathan Hurter for his time and contributions to the success of this paper.

## References

1. United States Marine Corps. Combat Hunter profiling B2E2965/B2E2967 [Student handout] (n.d.). [http://www.harisingh.com/Combat\\_Hunter\\_Profiling.pdf](http://www.harisingh.com/Combat_Hunter_Profiling.pdf)
2. Maraj, C.S., Hurter, J., Badillo-Urquiola, K.A., Lackey, S.J., Hudson, I.L.: Investigating performance of Kim's game for behavior cue detection. In: Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (IITSEC). NTSA, Arlington (2016)
3. Salcedo, J.N., Lackey, S.J., Maraj, C.: Impact of instructional strategies on workload, stress, and flow in simulation-based training for behavior cue analysis. In: Lackey, S., Shumaker, R. (eds.) VAMR 2016. LNCS, vol. 9740, pp. 184–195. Springer, Cham (2016). doi:[10.1007/978-3-319-39907-2\\_18](https://doi.org/10.1007/978-3-319-39907-2_18)

4. Hart, S.: NASA-task load index (NASA-TLX); 20 years later. In: Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, vol. 50, no. 9, pp. 904–908. Sage Publications, Los Angeles (2006)
5. Maraj, C., Hurter, J., Aubrey, W., Wolfe, E., Hudson, I.: Contrasting instructional strategies suited to a detection task: examining differences in subjective workload. In: Proceedings of the 2017 Human Computer Interaction International Virtual, Augmented, and Mixed Reality Conference. Springer (in press)
6. Ishihara, S.: Ishihara's Tests for Colour Deficiency. Kanehara Trading, Tokyo (2013)
7. Yanardag, M., Birkan, B., Yılmaz, İ., Konukman, F., Ağbuğa, B., Lieberman, L.: The effects of least-to-most prompting procedure in teaching basic tennis skills to children with autism. *Kinesiol. Int. J. Fundam. Appl. Kinesiol.* **43**(1), 44–55 (2011). <http://hrcak.srce.hr/69602>
8. Roh, Y.S., Issenberg, S.B., Chung, H.S., Kim, S.S., Lim, T.H.: A survey of nurses' perceived competence and educational needs in performing resuscitation. *J. Continuing Educ. Nurs.* **44**(5), 230–236 (2013). doi:[10.3928/00220124-20130301-83](https://doi.org/10.3928/00220124-20130301-83)
9. Schuster, D., Rivera, J., Sellers, B.C., Fiore, S.M., Jentsch, F.: Perceptual training for visual search. *Ergonomics* **56**(7), 1101–1115 (2013). doi:[10.1080/00140139.2013.79048180/00140139.2013.790481](https://doi.org/10.1080/00140139.2013.79048180/00140139.2013.790481)