

# A Holistic Evaluation of Task View Format for Training a Simulated Robot-Assisted EOD Task

James P. Bliss<sup>1</sup>✉, Eric T. Chancey<sup>1</sup>, Alexandra B. Proaps<sup>1</sup>, and Peter Crane<sup>2</sup>

<sup>1</sup> Old Dominion University, Norfolk, VA, USA  
{jbliss, echan004, aproaps}@odu.edu

<sup>2</sup> A2-T2, Inc., Orlando, FL, USA  
pcrane@a2-t2.com

**Abstract.** Recent decreases in psychophysiological recording method cost have enabled researchers to more easily supplement questionnaire and performance based indices of cognitive constructs like workload and situation awareness. The current paper describes the results of an experiment to compare single- and dual-view task interfaces for simulated unmanned vehicle navigation and object manipulation. Testing ten ROTC students in a within-condition experiment, researchers found that task completion time and eye gaze data revealed general learning trends and preferences for the dual-view condition. The results align with prior documentations of dual-view advantages and provide useful estimates of simulator learning speed.

**Keywords:** Robot · Simulation · View · Training · Control · FRAM

## 1 Introduction

Researchers have demonstrated that experts cognitively process information differently than novices. Early research by Chase and Simon [1] demonstrated that expert chess players could recall the position of pieces on a board more accurately than novices due to their advanced domain knowledge. With regard to unmanned vehicle control and monitoring, operators have been shown to verbalize aspects of the control task in different ways depending on their level of expertise [2].

The structure and delivery of task training often influences the rapidity and comprehensiveness of mastery. In some cases, unmanned vehicle operators may have complete classroom training with minimal operational practice. Others, however, may receive comparatively little classroom training but considerable practical experience. The variability of practice may be compounded by interfaces used for vehicle control. In particular, the visual displays used to depict task elements may facilitate knowledge building or may inhibit it.

With current interface evaluation methods, experts use the interface and fill in questionnaires many times to identify negative system traits. Unfortunately, novice users often display confusion and frustration, and commit errors when encountering embedded negative system traits. This outcome is especially true in reduced manning level environments where each individual has more and different types of tasks.

The purpose of this research was to use the Fused Realities Assessment Modules (FRAM) system to measure novice operator opinions and performances during a simulated unmanned vehicle operation scenario. By assessing physiological data in a low-cost fashion, FRAM represents a neuroergonomic approach for construct measurement [3]. FRAM scenarios reflect critical knowledge, skill and ability components and may allow comprehensive performance tracking. Specific neuroergonomic data sources are meant to enable real-time assessment and tracking of critical cognitive constructs such as mental workload and situation awareness.

For workload, pupil gaze has been used as a diagnostic indicator of workload in two ways. First, dwell time may be an indication that the individual is extracting information from certain areas. Dwell times differ by level of expertise, where novices dwell longer than experts as a function of workload. Second, in multitask environments fixation points indicate the task that is the greatest source of resource allocation and thus workload [4]. The reliance on eye tracking within FRAM should allow estimation of workload, especially in combination with assessment of task performance and data from a questionnaire such as the NASA-TLX [5].

With regard to situation awareness, there are several methods available to assess a user's environmental element perception (Level 1 SA), comprehension (Level 2 SA) and projection across time (Level 3 SA) [6]. Wickens & Hollands [4] have suggested that the construct of SA is domain specific; this frequently impacts the way in which it is measured. FRAM allows for estimation of workload by considering data from subjective questionnaires such as the Situation Awareness Rating Technique [7] as well as by examining eye tracking behavior and conventional task performance data such as reaction time and task success.

To determine the potential for FRAM to reflect the growth of task expertise in novices, we measured task speed and accuracy, operator workload, and situation awareness as operators interacted with a single-screen and a dual-screen display.

The use of methods and tools to assess and identify cognitive factors that impact human-technology performance must be carefully evaluated. For the current research, the FRAM tool leveraged a variety of neuroergonomic measures to supplement the use of task measures and traditional paper-and-pencil measures to comprehensively reflect human-technology performance. After consulting available published research, our team selected a focused range of the most pertinent cognitive components related to UGV operator performance: mental workload and situation awareness.

One purpose for this research was to demonstrate the utility of the FRAM tool suite for measuring cognitive constructs in real time during complex task performance. Standalone measures of mental workload and situation awareness have faced criticism because of questionable validity and challenges related to data collection [8, 9]. Recently, researchers have devoted effort successfully to triangulate measures for complex constructs such as workload [10]. The data reported here represent an initial attempt to demonstrate the ability of FRAM to accomplish such triangulation.

The second purpose for this investigation was to investigate the ability of FRAM to assess and track the growth of expertise within the context of simulated small UGV performance. This goal was approached with the specific intent of comparing task view format as an independent variable of interest. Prior researchers have demonstrated

performance gains and subjective preference for multiple screen views [11]. However, such research has not documented changes in such data across the task learning continuum. For this reason, we made no specific hypotheses regarding superiority of view format.

## 2 Method

### 2.1 Experimental Design

To ensure adequate statistical power with a relatively small sample of task experts, the experimenters employed a one-way, repeated-measures design. Screen view format (single-screen or dual-screen view) was manipulated within participants. Dependent measures included several performance based and neuroergonomic data streams within the FRAM suite. First, the experimenters measured task performance speed in seconds. Task accuracy was reflected by the number of charges successfully placed within a designated area. Neuroergonomic measures included eye tracking gaze time for particular screen regions (in seconds), electroencephalograph energy bands, and questionnaire data for cognitive workload (NASA-TLX) and situation awareness (SART). The experimental task required simulated UGV navigation and manipulation of a simulated detonator using an arm. Participants relied on a single-screen view to accomplish this in one session and a dual-screen view in a second session (see Figs. 1, 2 and 3).



**Fig. 1.** Experimental layout with task and monitoring stations



Fig. 2. Task interface – single screen view



Fig. 3. Task interface – dual screen view

## 2.2 Participants

Ten ROTC students from Old Dominion University participated for class credit (9 male, 1 female). Average age was 20.9 years (20-24). Eight of ten reported no prior experience using robots.

## 2.3 Materials

Questionnaires included the Informed Consent Form, Demographics Questionnaire, the SART situation awareness questionnaire [7], the NASA-TLX workload questionnaire [5], and a post-scenario questionnaire. The SART and the NASA-TLX have been widely used by researchers for decades and have demonstrated acceptable psychometric properties. Eye tracking was reflected as gaze duration in seconds within pre-defined areas of interest (task screen and birds-eye map). Task performance data included time (in seconds) to pick up a simulated detonator and total time to complete a navigation course with it.

The experimental task scenario was designed to resemble a common occurrence: using an unmanned ground vehicle to approach explosive charges, pick up the charges, and transport them to a safe location. The scenario was designed systematically through task analytic interviews with U.S. Marines and Army experts. That process included determining the behaviors, skills, and abilities necessary for completing the task. The research team also completed a comprehensive error analysis pertinent to the particular visual task interface used. Pilot testing of the scenario was accomplished by testing military expert personnel at Fort Benning, Georgia.

## 2.4 Procedure

After arriving at the laboratory, participants first completed the Informed Consent form and Demographics Questionnaire. The demographics questionnaire included items to determine participant sex, age, and experience with robotic devices. After completing the questionnaire, participants were randomly assigned to a task session sequence (single screen or dual screen first) and were trained to navigate the UGV within a practice virtual environment. After they indicated that they could effectively move the UGV in all directions, they were trained to use the simulated robotic arm to grasp and hold the simulated charges.

Following task training, participants were asked if they had any questions or needed clarification about the required task. Once participants' questions had been answered, they then completed two task sessions, separated by a five-minute break. Participants were free to take as much time as necessary to complete the task. During the break between sessions, participants completed the SART and the NASA-TLX questionnaires. Following the second session, participants completed an opinion questionnaire to describe the strategies they employed, their estimation of their own performance, and their preferences for each display condition. After completing the opinion questionnaire, participants were debriefed and dismissed.

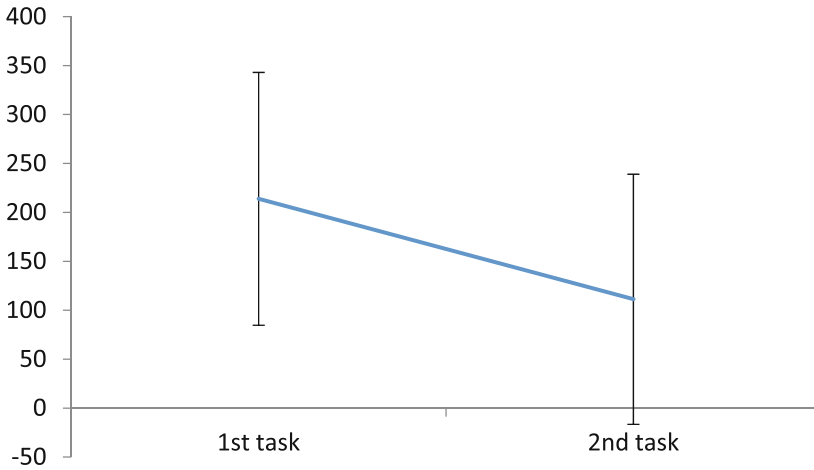
## 3 Results

Because of the novel nature of this research,  $p < .10$  was used as a significance criterion. For the analyses below, data from the one female participant were excluded because of difficulty acquiring EEG signals.

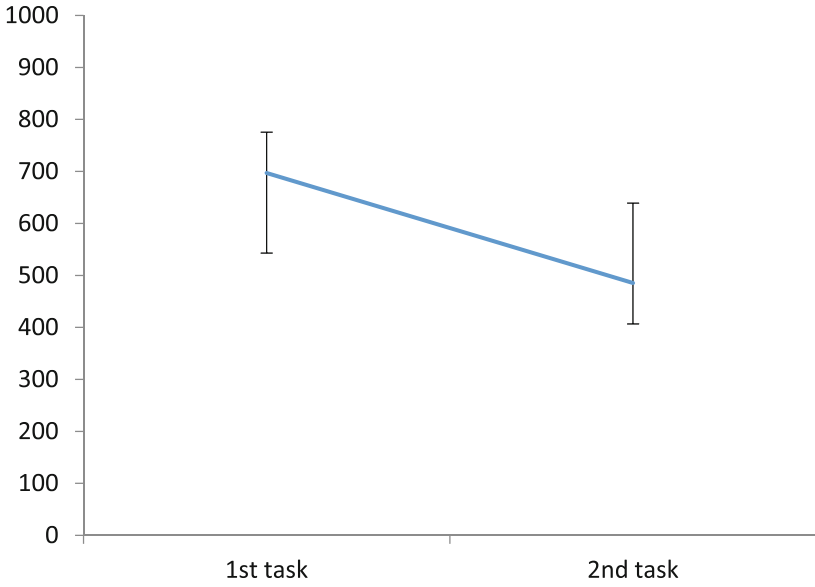
Few performance and subjective differences were evident between Single and Dual Screen conditions.

Participants did prefer the Dual Screen layout based on written and verbal comments. They perceived the dual screen layout to be less confusing than the single screen layout ( $p = .035$ ). The effort factor of the NASA-TLX approached significance, indicating that participants expended more effort to process task imagery on two screens ( $p = .107$ ). There were no significant differences found between single and dual screen conditions for situation awareness, time on task, time taken to pick up the C4 charge, and eye gaze duration ( $p > .05$ ).

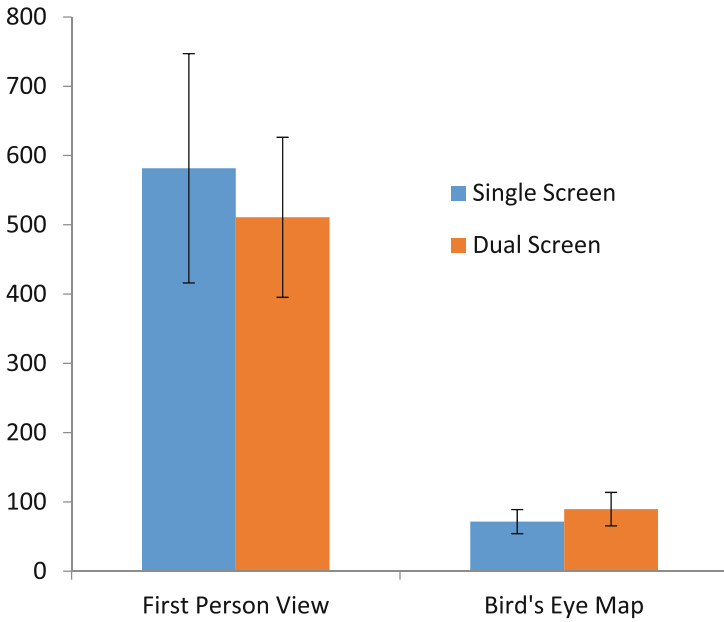
With regard to training, differences in time on task were significant ( $p = .06$ ) and time taken to pick up the C4 charge ( $p = .085$ ) between the first and second sessions, regardless of screen layout (less time required for the second task; see Figs. 4 and 5). It is also important to emphasize that the relative amount of time participants devoted to viewing the primary task view and the birds-eye map varied between the single- and dual-screen formats. Specifically, participants focused longer on the birds-eye map when completing the task in dual-screen format (see Fig. 6).



**Fig. 4.** Total time (in seconds) taken to pick up the C4 charge between first and second task (N=9). Note: Error bars indicate standard errors.



**Fig. 5.** Total time (in seconds) taken to complete task between first and second tasks (N= 9)



**Fig. 6.** Means and standard errors for gaze duration on First Person Shooter view and Bird's Eye Map view for Single and Dual screen layouts (N= 9).



## 4 Discussion

These data appear to demonstrate that the combination of performance, opinion, and physiological measures allow experimenters and designers to draw conclusions not possible with limited data sets. The data also provide an estimate of task improvement on a simulated unmanned vehicle task as a function of time and experience.

With regard to the data collected, the ability of the FRAM suite to support triangulation of measures allowed meaningful conclusions to be drawn about the single- and dual-screen comparator interface styles. Had the experimenters relied on only one source of data (for example, isolated physiological measures like heart rate or brain energy readings), the conclusions would likely have been confusing or even misleading. Using FRAM to delve deeper empowers investigators to isolate finer distinctions between conditions. These distinctions can be used to inform further research or training sessions.

Paradoxically, participants showed preference for the dual-screen task view, though it required more effort for them to process information using that layout. This result concurs with prior research related to computer productivity. As reported by Dell, Inc. [11], three commissioned studies (conducted at Wichita State University, Georgia Institute of Technology, and the University of Utah) presented the same conclusion: using multiple monitors to accomplish tasks led to faster task speed, greater task efficiency and higher user satisfaction. This finding echoes similar research conducted by the Microsoft Research team eight years earlier that concluded multiple monitors may enable a 9 to 50 % increase in task productivity [12]. It is important to point out that much of the early research demonstrating greater preference and productivity for a dual-display setup confounded number of screen views with overall task display size. However, in the current research, the overall size of the task view was equivalent; the difference was that the dual-screen condition included a top-down view of the task. By all accounts, participants concluded that the addition of this view led to a better ability to accomplish the object manipulation and transport aspects of the task.

The clear learning effect from Session 1 to Session 2 (regardless of which view layout occurred first) is important to consider. Researchers have for decades demonstrated that psychomotor task performance requires considerable time to master [13]. Early military research using virtual environment tasks echoed the early findings. Lampton, Knerr, Goldberg, Bliss, and Moshell [14] showed that participants required significant time to master even simple control movements when performing tasks in virtual environments. Participants tested in the current research clearly benefitted greatly from practice, though many often played games with the same control device used. One important lesson to be learned is that military personnel cannot be assumed to simply transfer prior knowledge with video games to unmanned vehicle control and object manipulation. Such skills require considerable practice, especially when the task view is divided among visual displays.

The current research demonstrates the utility of the FRAM tool suite for assessing and understanding complex tasks. Future researchers may benefit from examining further the FRAM method of neuroergonomic measurement to isolate the bounds of its utility.



## References

1. Chase, W.G., Simon, H.A.: The mind's eye in chess. In: Chase, W.G. (ed.) *Visual Information Processing*, pp. 215–281. Academic Press, New York (1973)
2. Andre, T., Hunt, J., Crews, A., Gluck, K.: Effect of fatigue and type of training strategy on UAV simulator performance. In: *Proceedings of the Human Factors of UAVs Workshop*. Mesa, AZ: Cognitive Engineering Research Institute (2007)
3. Parasuraman, R., Rizzo, M.: *Neuroergonomics: The brain at work*. Oxford University Press, London, UK (2008)
4. Wickens, C.D., Hollands, J.G.: *Engineering Psychology and Human Performance*, 3rd edn. Prentice Hall, Upper Saddle River, New Jersey (2000)
5. 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*. North Holland Press, Amsterdam (1988)
6. Endsley, M.: Toward a theory of situation awareness in dynamic systems. *Hum. Factors* **37**(1), 32–64 (1995)
7. Taylor, R.M.: Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In: *Proceedings of the AGARD AMP Symposium on Situational Awareness in Aerospace Operations, CP478*. Seuilly-sur Seine: NATO AGARD (1989)
8. Miller, S.: Literature review: Workload measures. (Final Report No. N01-006). Iowa City, IA: National Advanced Driving Simulator (2001)
9. Salmon, P., Stanton, N., Walker, G., Green, D.: Situation awareness measurement: A review of applicability for C4i environments. *J. Appl. Ergon.* **37**(2), 225–238 (2006)
10. Rusnock, C., Borghetti, B., McQuaid, I.: Objective-analytical measures of workload – the third pillar of workload triangulation? In: Schmorow, D.D., Fidopiastis, C.M. (eds.) *AC 2015*. LNCS, vol. 9183, pp. 124–135. Springer, Heidelberg (2015)
11. Dell, Inc. Dual monitors boost productivity, user satisfaction [online] (2011). [www.dell.com/downloads/global/products/monitors/en/dual\\_monitors\\_boost\\_productivity\\_whitepaper.pdf](http://www.dell.com/downloads/global/products/monitors/en/dual_monitors_boost_productivity_whitepaper.pdf), January 30, 2016
12. Ross, S.: Two screens are better than one [online] (2003). <http://research.microsoft.com/en-us/news/features/vibe.aspx>, January 30, 2016
13. Reynolds, B.: The effect of learning on the predictability of psychomotor performance. *J. Exp. Psychol.* **44**(3), 189–198 (1952)
14. Lampton, D.R., Knerr, B.W., Goldberg, S.L., Bliss, J.P., Moshell, M.J.: *The Virtual Environment Performance Assessment Battery (VEPAB): Development and evaluation*. (Technical Report No. 1029). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences (1993)