2022

Differential Workload Levels for Primary and Secondary Tasks using Virtual Reality

Ritesh Chandrasekaran
University of Central Florida

HONORS THESIS FOR DIFFERENTIAL WORKLOAD LEVELS FOR PRIMARY AND SECONDARY TASKS USING VIRTUAL REALITY

BY

RITESH CHANDRASEKARAN

A thesis submitted in partial fulfillment of the requirements for

the degree of Bachelor of Science

in the Department of Psychology

in the College of Sciences

at the University of Central Florida Orlando, Florida

Spring Term

2022

Thesis Chair: Dr. Pamela Wisniewski

TABLE OF CONTENTS

List of Figures

ABSTRACT

Virtual reality (VR) has modernized the way that training has been done across multiple domains. Combining the benefits of using VR with the ability to measure workload levels from eye tracking data is a promising area of opportunity for all disciplines and jobs that require hands-on training. With VR becoming of increasing interest for optimizing training simulations and protocols for work around the world, this study aims to determine differential workload levels in virtual reality by using objective performance data in a simulation and subjectively assessing workload through the NASA-TLX survey. For this study, a 3x1 repeated measures ANOVA was conducted with one control, primary task only, and two experimental conditions, dual tasking with secondary auditory and visual tasks. 49 particpants were recruited from the local UCF area to study how varying task objectives impacted performance on target identifications, eye fixation counts, and their NASA-TLX survey scores. By doing so, we were able to compare primary task performances that were done concurrently with secondary tasks and measured task interference as a result of workload. Overall, we found that if a secondary task pulls from the same sensory modality (e.g.. visual, auditory) as the primary task, then primary task performance is not significantly impacted. This finding can help future training programs be designed in such a way so that the user is not overburdened and can adequately complete the required tasks. Another key takeaway from the study was that fixation counts may not be a reliable measure of workload as in the dual visual condition, the fixation counts were at the highest but that was not reflected in the workload assessment for the NASA TLX score.

Chapter 1:Background

1.1 Virtual Reality and Workload Assessment

The use of Virtual Reality (VR) has grown tremendously since its creation. Since the 1950s, VR has grown to include head mounted displays and movement capability and now VR has grown and expanded into multiple domains (Boas, 2013). In the medical field, VR assisted surgeries look to increase the accuracy and precision of the surgery and decrease the overall amount of time that it takes to perform the surgery (Albani, 2007). In the manufacturing field, a VR teaching tutorial on how to assemble products or train new staff will decrease the amount of training time and increase the efficiency and output of the organization (Mekni & Lemieux 2014). In a military context, the enhanced realism and immersion looks to better train future military units by better equipping them for various different scenarios and situations (Siter, 2019).

Workload is “the perceived relationship between the amount of mental processing capability or resources and the amount required by the task” (Hart & Staveland 1988). Mental workload is most simply defined in terms of the allocation of processing resources to meet task demands (Matthews, 2015). Workload is an important area of study because it helps to show where an operator can perform at an optimum level and how many tasks can be done by the operator while still maintaining a highly efficient output. It is important to know how many tasks the operator can carry out at one time because it helps us design optimized training programs and allocate the appropriate amount of workload to the operator.

Workload was originally thought of as a unitary construct in the sense that there was only a singular large reservoir of information processing resources to deal with task demands (Matthews, 2015). However, when these task demands become varied, i.e. introducing a visual and auditory component into the tasks, dissociations between the subjective and performance based workloads measures may arise (Wickens, 1984). This seems to suggest that there is not a direct correlation between the cognitive resources present and the demands of the task. This would also suggest that workload is a multivariate construct with different modalities. This is best explained by Wickens' multiple resource theory, which stipulates that workload is different when different modalities are used (1980, 2002). As can be seen in Figure 1, what the Multiple Resource Model shows is that there is a resorvior of information processing resources that be accessed based on the input that is received, which in this case can be classified as visual, auditory, tactual, or olfactory. After the perception and engagement of a specific modality, the individual will process and determine an action that can be based on off of three reasoning techniques. This can be subconscious, symbolic, or lingustic depending upon the sensation and action that wil be required.

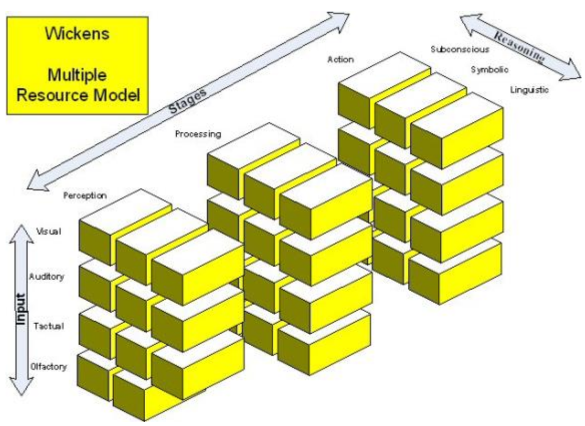


Figure 1 Wickens Multiple Resource Theory Model

The differential modalities of workload are the primary focus of this research. This theory is based on the principle of dual-tasking. Dual-tasking can be defined as two unique tasks being attended to simultaneously (Hazeltine et al., 2006; Koch et al., 2018). Multitasking, on the other hand, can be done in a sequential manner or a likewise concurrent manner. When a task is completed concurrently, there is continuous interference of one task over another as there is direct competition for information processing resources. According to Wickens, if two tasks demand resources from the same sensory modality, there will be performance degradation due to both tasks drawing from the same source of information processing resources. To elaborate on this idea, there will be more task interference if the tasks at hand are two visual tasks, rather than one visual task and one auditory task.

.Another related but different theoretical model for workload is Lavie's attentional model (Lavie, 2004). It focuses on interference by concurrently occurring stimuli that come from two different modalities. One of the key points that is raised is the centered versus peripheral visual processing. Centered visual processing is the primary visual task that most visual information processing resources are directed towards while peripheral visual processing is the secondary visual task that is interfering with the primary tasks and competing for processing resources. Could operators get distracted due to the nature of the study and the study looks to ask if focusing on task-relevant things prevent even task-irrelevant stimuli from getting perceived? Or does focusing on task-relevant things help from getting behaviorally distracted from or even remember task-irrelevant stimuli? Task-relevant distractors are stimuli that appear that deviate from the primary task but are still associated with the primary task. Task-irrelevant distractors are stimuli that are non-associated with the primary task and serve to distract attention away from the primary task. The results from Lavie's research indicate that distractor perception can be prevented when attention to task-relevant stimuli involves a high cognitive load and while distractors may be perceived they are filtered out depending on the type of load like working memory. The conclusions that were drawn from Lavie's study can be applied so that people can achieve goal-orientated tasks with very minimal intrusion from the task-irrelevant stimuli (Lavie, 2004).

The overall considerations of the stated theories and real-world applications leads to the postulation of the hypothesis that states primary task performance under dual-task conditions will be lower than primary task performance under single-task conditions. The scope of this paper will focus on the specific application of VR and eye-tracking to assess workload modalities specific to the military. However, our findings may be generalizable to other fields, such as the medical and transportation industries. In particular, there are military applications with Integrated Visual Augmentation System (IVAS) that the military is looking to implement (IVAS, 2018). The IVAS system emphasizes real-time tracking and is expected to monitor the state of the soldier where the workload is the state. There will be an increase in visual acuity training and the military can prepare proper courses to optimize soldier performance. IVAS intends to help provide individual feedback on soldiers' performances in mission rehearsals to help increase performance on the battlefield. A potential downside of such a system is that there still is a limited field of view providing a challenge of poor generalization to the real world where peripheral vision could make a significant the difference.

1.2 Use of Virtual Reality in Training Simulations

One of the added benefits of VR is the immersion, increased realism, and training efficiency that can be done in a military setting. It will prove cost effective and more efficient to help simulate real world environments (Rathnayake, 2018). The disadvantages include low external validity and not accurate generalizations to the real world in terms of the variable environments that will be encountered. (Cronbach & Meehl, 1955; Noack et al., 2014). However, ultimately the cons are outweighed by the pros and VR is being used in the military to increase operational capabilities like aided target recognitions, inclusion of AI software, and affordance of eye tracking metrics on the move.

1.3 Workload Measurement

The purpose of workload measurements is to provide an idea of where an operator is more prone to error or overload (Matthews 2019; Viduilch and Tsang 2012). The operator can only process a limited amount of information at once. The metrics that may be involved include subjective scales like the NASA-TLX survey (Matthews, 2015), performance-based, which is data driven, and psych physiological like eye-tracking metrics (Hancock and Chignell 1988, O'Donnell and Eggemeier, 1986; Viduilch and Tsang).

Subjective scales like the NASA-TLX survey have been critiqued in the past for having an inherent limitation of being solely based on the perceived experience of the task performer. (Dekker and Nyce 2015) (de Winter, 2014; Matthews et al., 2019). However, the NASA-TLX survey is still useful in determining facets of the multidimensional construct of workload. It also helps to predict performances and is an operational use of the measure; subjective measures are sensitive to measuring workload to the number of tasks being performed and provide insight into self-regulatory strategies (de Winter, 2014, Abich et al., 2013; Abich et al., 2017).

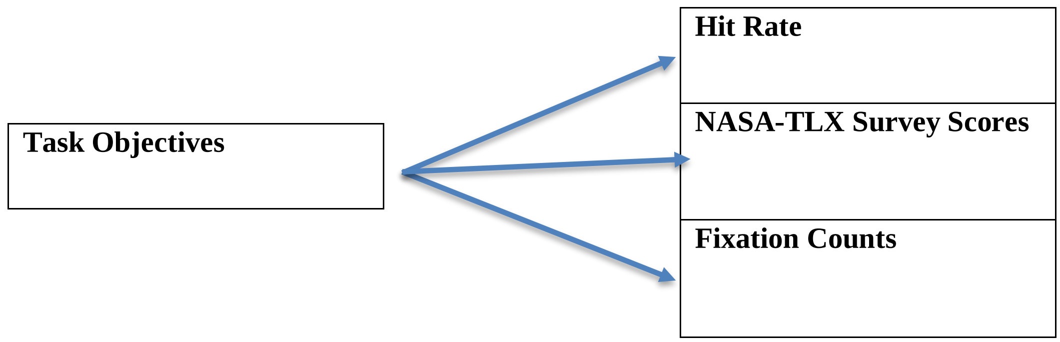
Chapter 2: Theoretical Model and Experimental Design

2.1 Purpose of the Study

The purpose was to determine differential levels of workload in VR, comparing performance and subjective workload.

2.2 Theoretical Model and Variables Explained

Independent Variables ***Dependent Variables***



This theoretical model is set up so that a participant will run through the 3 different task modalities for the 3 different levels that are associated with the independent variable. The 3 levels are as follows: single primary visual task which will act as the control group, dual tasking of the primary task with a secondary visual task, and dual tasking of the primary task with a secondary auditory task. As for the dependent variables they are operationalized so that they are a measure of workload. The hit rate will be how many insurgents they were able to identify correctly out of how many insurgents were actually in the simulation. After the simulation is complete, they will take the NASA-TLX survey to provide their own subjective metric of their workload. The same principles are applied with the visual and auditory task which are gone into detail below with the primary task occurring concurrently. The hit rate in the visual task is the number of times that they correctly identified if a target marker is present on their minimap out of the number of total target markers present. The hit rate in the auditory task will be the number of times that they correctly answer the question posed out of the total number of questions.

2.3 Description of each Task

Threat detection: People would walk across the screen in a horizontal line and the task is to identify people with an AK-47 as a threat. There are 226 total characters and 30 total threats giving an event rate of 45 characters per minute and a threat probability of about 13%.

Visual: A given structure was described to the participant and the target color was given before the scenario started. The participant would then press yes or no on the VR remote to indicate if the target marker of the appropriate color was present. The order of the target color was also randomized and counterbalanced. The refresh rate for the stimuli was approximately every 5 seconds.

Audio: In the VR headset, a predetermined script was read to the participant asking simple logic questions and the participant responded in a similar yes or no fashion using the VR remote. The same three scripts were used and they were not randomized.

2.4 Performance measures

Threat detection task: Hit rate is a proportion of correct threat identified over the total number of threats. A false alarm rate is a proportion of incorrect threats identified over the total number of threats. Misses are a proportion of missed threats over total number of threats.

Visual tasks: The correct answer rate is a proportion of the number of target markers that are correctly identified as being present or absent over the total number of target markers that appeared.

Audio tasks: The correct rate is a proportion of the number of audio questions that are correctly answered yes or no over the total number of audio questions.

2.4.1 Subjective workload measures

Workload was also measured through a subjective scale using the NASA-TLX survey after each run was completed. The survey measures different components of workload like mental demand, physical demand, temporal demand, effort, frustration and performance and gives the user the scale of low to high with the exception of performance which is graded on a scale of good and poor. To assess the scales objectively, after the position on the scale is chosen by the participant, each subsection has its own raw score. Then each raw score was multiplied by the scale score for each category and then divided by 15 to calculate the overall workload index to get an indication of how much work the participant felt that they were putting into the task scenario.

2.5 Statement of Hypotheses

Based on Lavie's attentional model and Wickens Workload model where task degradation will occur due to the competition of cognition resources as a result of the introduction of the secondary task the following hypotheses are given (Wickens, 2002).

Hypothesis	**Statement**
H1	Hit rate for task completion will be the highest in the single primary task condition.
H2	Hit rate for task completion will be higher in the dual visual task

	condition than dual auditory task condition due to the visual task sharing the same sensory modality as the primary task.
H3	NASA TLX scores will be the highest in the dual auditory task.
H4	NASA TLX scores will be higher in the dual visual task than single primary task due to increased perceived workload.
H5	Fixation counts will be the highest in the dual visual task condition.
H6	Fixation counts will be higher in the dual auditory task than the single primary task due to increased task demands.

Chapter 3: Methods

3.1 Study Procedure

First, consent to participate in the research study, which was IRB approved, was obtained using different forms, which were dependent on which sign-up system was used by the participant. After consent was obtained, the participants were screened using the Ishara Color Vision Test to ensure that each participant had adequate color vision. They also answered some demographic and eligibility questions. Then the participant s completed a self-guided PowerPoint Training where the different tasks were explained in depth. They were also taught how to don the VR headset and how to use the remote. The three scenarios that were described in the PowerPoint were a single threat detection task, an audio task, and a visual task. After each task was described a small checkpoint was used to ensure that the participant understood the task well enough to perform it. Between each participant, the order of which the tasks were shown were randomized and counterbalanced to help control bias. Once the training was complete, the participant completed surveys to rate levels of stress and usability. After this, the participant went through each scenario that lasted about 15 minutes. They completed surveys after each scenario. The total study duration was 2.5 hours. The minimum range for the study was about 90 minutes and the maximum range was around 200 minutes with the variability coming from how long each participant took for the training task and to complete the associated surveys. The participants were either compensated with money or with volunteer hours, depending on which they chose.

3.2 Materials

The materials that were used in this study included the HTC Pro Eye Vive Headset and Remotes, the Task Completion training slide deck, the simulations were created in Unreal Engine Gaming

Software, eye-tracking was done through the use of the Ranipal Eye-Tracking Software, and finally the NASA-TLX Survey was used to assess the participants workload levels.

3.3 Data Analysis Approach

The design for this experiment will be following repeated measures ANOVA design because the variability in the conditions is reduced and the only variability that needs to be accounted for is the variability withing subjects with 3 different levels. An ANOVA is also preferred to performing multiple t-tests because when running multiple t-tests the chance of making a type 1 error increases whereas in an ANOVA the chance of a type 1 error is controlled to be at 5% (Fields, 2015).

3.4 Participant Recruitment and Demographics

49 people were recruited using a variety of methods which include cash incentives, volunteer hours or course credit. The amount of participants required to carry out a statistically significant experiment was determined by using a G-power analysis test. The participants are going to be recruited through the use of the IST and UCF SONA systems as well as reaching out the local ROTC unit. Participants must be 18 years of age or older, adequate color vision, not be pregnant, have not ingested alcohol, antihistamines, decongestants or sedatives 24 hours prior to the study, not have a history of epilepsy, seizures or motion sickness, or finally not have ingested caffeine to the equivalent of 2 cups of coffee 24 hours before the study. If a participant does not meet any of the criteria, there is sufficient grounds for them to not participate in the study.

For this experiment, the mean age of participants was 20.28 years and 20 female participants and 29 male participants. 37 of the participants reported that they had less than 4 years of college education, 11 participants reported that they completed 4 years of college, and 1 participant did not respond to the question. The self-reported average of the number of hours spent on video game usage was 3.67 hours. Only one participant self-reported military experience in which they served three years.

Chapter 4: Results

4.1 Descriptive Statistics and Correlation Matrices

Descriptive Statistics			
	Mean	Std. Deviation	N
HitRate_P	.8948	.11594	36
HitRate_V	.8503	.13897	36
HitRate_A	.8170	.15073	36
TLX_Global_P	34.2029	17.11342	46
TLX_Global_V	44.1667	15.05135	46
TLX_Global_A	53.8949	14.92166	46
Fixation_Count_P	572.2619	43.13844	28
Fixation_Count_V	665.2143	46.37441	28
Fixation_Count_A	562.2262	59.32501	28

The following descriptive statistics were given for the conditions and levels conducted in the study with P referring to the single primary task condition, V referring to the dual visual task condition, and A referring to the dual auditory task condition.

Correlations

		HitRate_P	Fixation_Count_P	TLX_Global_P
HitRate_P	Pearson Correlation	1	.056	-.178
	Sig. (2-tailed)		.737	.259
	N	43	38	42
Fixation_Count_P	Pearson Correlation	.056	1	-.144
	Sig. (2-tailed)	.737		.394
	N	38	38	37
TLX_Global_P	Pearson Correlation	-.178	-.144	1
	Sig. (2-tailed)	.259	.394	
	N	42	37	46

Correlations

		HitRate_V	Fixation_Count_V	TLX_Global_V
HitRate_V	Pearson Correlation	1	.410*	-.247
	Sig. (2-tailed)		.014	.115
	N	44	35	42
Fixation_Count_V	Pearson Correlation	.410*	1	-.372*
	Sig. (2-tailed)	.014		.030
	N	35	35	34
TLX_Global_V	Pearson Correlation	-.247	-.372*	1
	Sig. (2-tailed)	.115	.030	
	N	42	34	46

*. Correlation is significant at the 0.05 level (2-tailed).

Correlations

		HitRate_A	Fixation_Count_A	TLX_Global_A
HitRate_A	Pearson Correlation	1	-.013	-.438**
	Sig. (2-tailed)		.938	.006
	N	40	36	38
Fixation_Count_A	Pearson Correlation	-.013	1	-.082
	Sig. (2-tailed)	.938		.639
	N	36	37	35
TLX_Global_A	Pearson Correlation	-.438**	-.082	1
	Sig. (2-tailed)	.006	.639	
	N	38	35	46

**. Correlation is significant at the 0.01 level (2-tailed).

4.2 ANOVA results for Hit Rate

A repeated-measures ANOVA determined that mean primary task hit rates differed significantly across three workload conditions ($F(2, 70) = 7.196$, $p = .001$). Maulchly's Test, $\chi^2(2) = 2.829$, $p = 0.243$ did not indicate any violation of sphericity. A post hoc pairwise comparison using the Bonferroni correction showed an increased mean hit rate proportion between the cognitive load condition with the secondary auditory task and the perceptual load condition with the secondary visual task (0.817 vs 0.850, respectively), but this was not statistically significant ($p = .435$). However, the increase in mean hit rate proportion did reach significance when comparing the perceptual load condition and cognitive load condition to solely the primary task condition (0.850 vs 0.897, $p = .047$) and (0.817 vs 0.897, $p= .003$). Therefore, we can conclude that the results for the ANOVA indicate a significant effect for the high cognitive load on hit rate which indicates that workload was increased as a result of this task condition.

4.3 ANOVA results for Fixation Counts

A repeated-measures ANOVA determined that mean fixation counts differed significantly across three workload conditions (F(1.428, 38.566) = 71.926, p = < .001). Maulchly's Test, $\chi^2(2)$ = 13.290, p = 0.001 did indicate a violation of sphericity, so the multivariate test of Greenhouse-Geisser was used instead. A post hoc pairwise comparison using the Bonferroni correction showed an increased mean fixation counts between the primary task condition and the perceptual load condition with the secondary visual task (572.262 vs 665.214, respectively), and this was statistically significant (p = < .001) However, there was a decrease in mean fixation counts between the primary task condition and the cognitive load condition which did not reach significance (572.262 vs 562.226, p = 1.000). When comparing the perceptual load condition to the cognitive load condition, there was a significant decrease in the mean fixation counts (665.214 vs 562.226, p= < .001). Therefore, we can conclude that the results for the ANOVA indicate a significant effect for the secondary task effects on eye-tracking as measured by the total amount of fixation counts.

4.4 ANOVA results for NASA-TLX scores

A repeated-measures ANOVA determined that NASA TLX scores differed significantly across three workload conditions (F(1.633, 73.506) = 41.580, p = < .001). Maulchly's Test, $\chi^2(2)$ = 11.180, p = 0.004 did indicate a violation of sphericity, so the multivariate test of Greenhouse-Geisser was used instead. A post hoc pairwise comparison using the Bonferroni correction showed an increased NASA TLX score between the primary task condition and the perceptual load condition with the secondary visual task (34.203 vs 44.167, respectively), and this was statistically significant (p = < .001) There was an increase in NASA TLX scores between the primary task condition and the cognitive load condition which did reach significance (34.203 vs 53.895, p = < .001). When comparing the perceptual load condition to the cognitive load condition, there was an increase in the NASA TLX scores which did reach significance (44.167 vs 53.895), p= < .001). Therefore, we can conclude that the results for the ANOVA indicate a significant effect for the secondary task effects on subjective workload measurements as measured by the score of the NASA TLX survey.

Chapter 5: Discussion

5.1 Discussion of Hit Rate Results

The topic of differential workload levels for primary task performances and the effect of secondary task interference in virtual reality was able to shed some light on workload levels and provide evidence for an answer to the question of if secondary tasks that drew from the same sensory modality as the primary task would impact performance. This will help to optimize training simulations for the army so that soldiers are able to increase their situational awareness and be better prepared for the field. When looking at the results of hit rate, in accordance with Wickens' Multiple Resource Theory (2002), hit rate success was the highest in the single task conditions, then the dual visual task, and finally the lowest in the dual auditory task. One of the tenets of the multiple resource model detail that the demand of the tasks themselves indicate how many information processing resources can be dedicated to the task completion. In the high cognitive load condition, the engagement of the auditory sensory modality, it required both working memory and cognitive analysis which was much more demanding than a visual search in the other conditions. This is where the task interference in the high cognitive load condition originated as it demanded more from the user compared to the other conditions which is why the hit rate responses congruently decreased in this condition.

5.2 Discussion of NASA-TLX results

When looking at the data that was given through the NASA-TLX, it showed that participants felt an increase in perceived workload the most in the dual auditory task condition. This can be attributed to the increase in workload management as noted by the subjective measurement of workload through the NASA-TLX survey. That is, because of the increased stress levels that are introduced by the secondary task, the primary task performance suffered. An analogous situation would be to watch one TV show and listen to another TV show concurrently, with the primary task being describing what you see while the secondary task is to recall what was said in the second show. This was seen by the introduction of the novel sensory modality in combining both auditory and visual input.

5.3 Discussion of Fixation Counts results

When looking at the data that was given through fixation counts, it is interesting to mention that from the correlation matrix, it can be seen that fixation counts and the NASA-TLX scores were negatively significantly related in the dual visual condition. Although, it is expected that the fixation counts in a visual task would increase due to the amount of visual information that needs to be processed, however the perceived workload should increase not decrease with this finding. This gives an indication that perhaps fixation counts and eye tracking metrics may not be a good proxy measure for real-time workload assessment and needs further research to explore the relationship there.

5.4 Limitations and Future Research

Some of the study limitations include technical errors like the VR headset dying during the trials, training PowerPoints crashing during the training part of the trial, and the loss of calibration of the VR remotes, which lead to misfiring during the trial, which all lead to a varying amount of participant data for each condition, however, there was sufficient data collected to exceed the 25 minimum participant requirement. To eliminate the technical errors, the latest model of VR headset technology should be used with the latest compatible software to run the simulation as well as a stronger internet connection so that training slides do not crash. The usage of the NASA-TLX survey is based on a subjective assessment of workload, however, self-reported data merits a discussion of recall bias and may not reveal a true indication of workload. This is why the NASA-TLX was used in conjunction with the objective eye-tracking data to indicate an accurate representation of the participant’s workload levels. The use of VR will also stand to be questioned as it is not a perfect representation of field experience for the participants; however, it can simulate an immersive reality and evoke similar responses to those that will be seen in the field.

As far as limitations are considered in the study population, a majority of the participants were recruited from UCF Psychology classes, but a minority did come from the UCF ROTC, which is a representative population of the soldiers that will use this research in the future. All participants went through a pre-experiment checklist to ensure they got the proper amount of sleep, they were of age, they were not pregnant, etc. Although these responses were based on self-reporting, to control this in the future, testing can be done to ensure that the participants are truthful, however this method may require more time and funding. It is understood that there are limits to this as testing criteria like pregnancy is invasive and requires the honesty and truthfulness of the participants to report this back.

Further research efforts are needed to create an accurate simulation that can train military cadets for field experience as well as monitor workload levels through eye-tracking. Overall, the results of this study support the need for a large-scale randomized controlled trial with actual military participants. This trial would also help to contain more bias in the experiment and also will grant more control to the experimenters so that more precise and accurate conclusions can be drawn from due to the proposed experimental nature of the new study. It is likely with the growing technology in virtual reality as well as growing costs for training and educating a new generation of soldiers will require the optimal usage of funds and resources. With this, researchers will be able to create evidence-based training simulations that include close to all variables that a soldier might experience in the field. For example, a new simulation may involve movement as well as a 360-degree target screen, or perhaps they may want to introduce other secondary tasks that draw from novel sensory modalities like a tactile modality that would require the discernment between two different objects based on texture. Overall, the key takeaways from this study are that workload increases in the dual-task condition and performance on primary tasks degraded the most when the secondary task posited the usage of a different sensory modality than the primary task. This is important to future training because training programs that utilize VR need to understand that concurrent task demands should demand similar goals. For example, in the healthcare field, VR-assisted surgeries should try to overlay all of the information visually rather than providing both visual and auditory information. This way performance will be optimized and results will be beneficial to all of the affected parties. If these considerations are not taken into account then the task will not be achieved to its fullest capacity.

In conclusion, when designing future tasks in a dual task training environment, creators should aim to have similar task demands that pull from the same sensory modality to prevent task interference, while future research on this topic can focus on the relationship between eye-tracking as a proxy measure for workload assessment.

References

Albani, J. M., & Lee, D. I. (2007). Virtual reality-assisted robotic surgery simulation. *Journal of Endourology*, *21*(3), 285-287.

Boas, Y. A. G. V. (2013, August). Overview of virtual reality technologies. In *Interactive Multimedia Conference* (Vol. 2013).

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281.

de Winter, J. C. F. (2014). Controversy in human factors constructs and the explosive use of the NASA-TLX: A measurement perspective. *Cognition, Technology & Work*, *16*(3), 289–297. https://doi.org/10.1007/s10111-014-0275-1

Field, A. (2014). *Discovering statistics using Ibm Spss statistics +Spss version 22.0*. Sage Publications.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology:* (Vol. 52, pp. 139–183). Elsevier.

Jansen RJ, van Egmond R, de Ridder H (2016) Task Prioritization in Dual-Tasking: Instructions versus Preferences. PLOS ONE 11(7): e0158511. https://doi.org/10.1371/journal.pone.0158511

Kaplan, A. D., Cruit, J., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (2020). The Effects of Virtual Reality, Augmented Reality, and Mixed Reality as Training Enhancement Methods: A Meta-Analysis. *Human Factors*, 0018720820904229. https://doi.org/10.1177/0018720820904229

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 451–468. https://doi.org/10.1037/0096-1523.21.3.451

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load Theory of Selective Attention and Cognitive Control. *Journal of Experimental Psychology: General*, *133*(3), 339–354. https://doi.org/10.1037/0096-3445.133.3.339

Matthews, G., de Winter, J., & Hancock, P. A. (2019). What do subjective workload scales really measure? Operational and representational solutions to divergence of workload measures. *Theoretical Issues in Ergonomics Science*, 1–31. https://doi.org/10.1080/1463922X.2018.1547459

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, *77*(12), 1321–1329.

Noack, H., Lövdén, M., & Schmiedek, F. (2014). On the validity and generality of transfer effects in cognitive training research. *Psychological Research*, *78*(6), 773–789. https://doi.org/10.1007/s00426-014-0564-6

Rathnayake, W. G. R. M. P. S. (2018). Usage of Mixed Reality for Military Simulations. *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, 1–5. https://doi.org/10.1109/ICCTCT.2018.8550993

Siter, B. (2019). Soldiers test new IVAS technology, capabilities with hand-on exercises. *US Army*.

Stanney, K., Fidopiastis, C., & Foster, L. (2020). Virtual Reality Is Sexist: But It Does Not Have to Be. *Frontiers in Robotics and AI*, *7*. https://doi.org/10.3389/frobt.2020.00004

Stanney, K. M., Mollaghasemi, M., Reeves, L., Breaux, R., & Graeber, D. A. (2003). Usability engineering of virtual environments (VEs): Identifying multiple criteria that drive effective VE system design. *International Journal of Human-Computer Studies*, *58*(4), 447–481.

Wickens, C. D. (1980). The structure of attentional resources. In R. S. Nickerson (Ed.), *Attention and performance VIII* (Vol. 8, pp. 239–257). Lawrence Erlbaum Associates.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177.