AN APPROACH TO ACCELERATED LEARNING AND PSYCHOPHYSIOLOGICAL MEASURES OF ENGAGEMENT

Anna L. Oskorus, Terence S. Andre,
Tiffany R. Ripley, Ryan E. Meyer
TiER1 Performance Solutions
Covington, KY
a.oskorus@tier1perormance.com,
t.andre@tier1performance.com,
t.ripley@tier1performanc.com
r.meyer@tier1performance.com

Cali M. Fidopiastis
University of Alabama at Birmingham
Birmingham, AB
cfidopia@uab.edu
Dee H. Andrews, Patricia C. Fitzgerald,
Air Force Research Laboratory
Mesa, AZ
dee.andrews@mesa.afmc.af.mil,
patricia.fitzgerald@mesa.afmc.af.mil

ABSTRACT

The government, education and business sectors have a common and critical need to dramatically accelerate the transition of employees from novice to expert and improve human performance. Accelerated learning and learner engagement are two areas of interest with few conclusive reports but great potential to drive business results. Recent investigations within military-relevant training environments suggest that electroencephalography (EEG) signatures of attention, memory, and workload can be validly assessed during learning and linked to learner engagement. The concept behind these measures is to track—in real time—perceptual, attentional, and cognitive workload states that would indicate increased learner engagement and performance. As part of a research project with the Air Force Research Laboratory, TiER1 Performance Solutions investigated principles of learning acceleration in the context of a training course for teaching supervisors how to detect and respond to cyber insider threats. This paper reports on two recent experiments conducted to investigate the design of TiER1's accelerated learning approach and improve our product. The first study compares the TiER1 accelerated training system (e.g., custom learning pathways, quizzes, game-based interaction, and feedback) to conventional multimedia training, while the second evaluates using objective psychophysiological measures to detect cognitive state changes of learners who receive the accelerated training system versus those who train using equivalent modules of conventional multimedia training. Results from the first study showed very experienced learners could test out of content they already knew. However, post-test assessment scores showed no significant differences between the two groups. In the second study, psychophysiological measures relating to cognitive workload were better able to demonstrate differences between the traditional and the game-based training material than performance measures alone. These objective measures of cognitive workload were also supported by subjective report data (e.g., NASA-TLX), further demonstrating the utility of using such measures when evaluating serious game design.

ABOUT THE AUTHORS

Anna L. Oskorus is a Senior Consultant for TiER1. She designs and manages the development of blended learning solutions for commercial, non-profit, and government organizations. She is the project manager and lead instructional designer for Phase II of the Accelerated Learning for Cyber Insider Threat Reduction project. She received her master's degree in Instructional Systems from Penn State University with an emphasis in interactive learning technologies.

Dr. Terence S. Andre is a Managing Director at TiER1 Performance Solutions where he directs the overall research and innovation for the company. He is an expert in human-computer interaction (HCI) with over 20 years experience optimizing human and system performance through the use of systems engineering and human factors design principles. Prior to joining TiER1, Dr. Andre was an Associate Professor at the United States Air Force Academy where he retired as a Lt Colonel. Dr. Andre holds a Ph.D. in Industrial and Systems Engineering from Virginia Tech, an M.S. in Industrial Engineering from Cal Poly, and a B.S. in Behavioral Sciences from the United States Air Force Academy.

Tiffany R. Ripley is an Associate Consultant at TiER1 Performance Solutions and works on many engagements that have an applied research component and human performance improvement element. She is a project manager, training specialist and researcher. She seeks to improve areas of workforce development and assists with organizational change efforts. In January 2010, Tiffany joined the Learning and Communications team at TiER1 and serves a diverse set of clients both in the government sector and commercial businesses including the U.S. Air Force, National Science Foundation, Luxottica, LexisNexis, Macy's, Futurestep, and Amerisource Bergen. Tiffany brings a breadth of research methodology practices and training experience to TiER1.

Ryan E. Meyer is a Senior Consultant at TiER1. Ryan has several years experience implementing technical solutions for web-based instructional content. His technical background includes desktop application development in Java, 3D modeling and simulation engines, web application development in PHP/MySQL, and web interface development in Flash and AJAX. Ryan has a BS in Computer Engineering with a minor in Mathematics from the University of Cincinnati.

Dr. Cali M. Fidopiastis is the Director of the Interactive Simulation (*i*Sim) Laboratory located at the School of Health Professions on the campus of the University of Alabama Birmingham (UAB). Her research focuses on understanding brain plasticity initiated during learning as individuals perform tasks in naturalistic environments. She is a specialist in the use of psychophysiological sensors such as near infrared spectroscopy (NIRs) and EEG to quantify emotion, effort, and cognitive processing in operational environments. The goal of Dr. Fidopiastis' work is to combine the understanding of head-mounted displays, computer graphics, and user perception to create optimal simulated and real-world training solutions for military, surgical, and rehabilitation applications.

Dr. Dee H. Andrews is the Directorate Senior Scientist (ST) for the Human Effectiveness Directorate of the Air Force Research Laboratory (AFRL). As a Senior Scientist (ST), Dr. Andrews is the AFRL's principal scientific authority for training research. Dr. Andrews' responsibilities include sustaining technological superiority for training by planning and conducting theoretical and experimental studies. He also is responsible for mentoring and developing dedicated technical staff to assure quality in training research, and he represents AFRL in training research matters to the external scientific and technical community.

Ms. Patricia Fitzgerald is a Research Psychologist at the Air Force Research Laboratory Warfighter Readiness Research Division in Mesa, AZ. Ms. Fitzgerald is conducting in-house empirical research and providing technical guidance on contracted programs to assess the effects of various instructional strategies for improving the effectiveness of flight instructor and cyber training. She served as Program Manager for the Improved Performance Research Integration Tool (IMPRINT) Training Algorithm Enhancement program. Ms. Fitzgerald holds a Master of Science degree in Aviation Human Factors and a Bachelor of Arts degree in Psychology. She previously spent 22 years working as a computer systems analyst and software engineer.

AN APPROACH TO ACCELERATED LEARNING AND PSYCHOPHYSIOLOGICAL MEASURES OF ENGAGEMENT

Anna L. Oskorus, Terence S. Andre, Tiffany R. Ripley, Ryan E. Meyer
TiER1 Performance Solutions
Covington, KY
a.oskorus@tier1perormance.com, t.andre@tier1performance.com, t.ripley@tier1performanc.com r.meyer@tier1performance.com

Cali M. Fidopiastis
University of Alabama at Birmingham
Birmingham, AB
cfidopia@uab.edu
Dee H. Andrews, Patricia C. Fitzgerald,
Air Force Research Laboratory
Mesa, AZ
dee.andrews@mesa.afmc.af.mil, patricia.fitzgerald@mesa.afmc.af.mil

INTRODUCTION

Training content is ever expanding. The number of skills and breadth of knowledge that military personnel are required to learn grows each day. Great benefit would occur if we could develop methods to accelerate learning. Accelerated learning is a major goal in almost every part of the work force (Moe, 2009; Jacobson, 2005; Office of Management and Budget, 2005). In the military, in fact, the need "to dramatically accelerate the transition from novice to expert in key military tasks" is seen as absolutely crucial to the continued effectiveness of U.S. fighting forces around the globe (Department of Defense, 2009). This paper describes a project aimed at investigating a research-based approach to speed up learning and to increase overall learner effectiveness through a game-based learning environment. The domain chosen to apply and evaluate these acceleration techniques was cyber insider threat. Specifically, we wanted to try to accelerate the learning of workforce supervisors so that they could potentially mitigate the threat of an employee intentionally stealing information or harming cyber networks. These types of skills are in critical need among supervisors in the workforce, yet there is very little opportunity to practice and gain experience. Game-based learning environments provide a possible approach to train these skills and support higher learning performance.

The Department of Defense (DoD) currently does not have a systematic plan for training its personnel about cyber insider threat. The Verizon Risk Team summarized the source of cyber breaches into the following categories:

- 92% from external sources
- 17% from insiders (approx. 100 breeches)
- < 1% implicated business partners
- 9 % involved multiple parties

These statistics represent 7 years of data, 1700+ breaches, and over 900 million compromised records (Verizon, 2011). The Verizon report provides a sobering picture of the seriousness of cyber insider threat in the business world, however we know of no such accounting that has been made in the DoD.

External threats to cyber networks receive the most attention, but malicious abuse of network access by employees can be even more devastating. Whereas cyber threats are easily identified through technical solutions, it is much harder to identify employees who intend to commit an insider crime. Supervisors need training to help them recognize the behavioral signals that an insider threat exists.

As part of a research project with the Air Force Research Laboratory, TiER1 Performance Solutions investigated principles of learning acceleration in the context of a training course designed to teach supervisors how to detect and respond to cyber insider threats. We developed a training system known as Accelerated Learning for Cyber Insider Threat Reduction (XL-CITR) and used theories from accelerated learning and learner engagement to evaluate this system for both DoD and commercial use.

Supervisors were selected as the population of interest for this training for two important reasons. First, interviews with supervisors and personnel specialists in the Air Force revealed that little, if any, of their training addressed the identification of the psychosocial behaviors that may alert them to potential malicious insider activity. The closest knowledge base that was identified was dealing with disgruntled or difficult people. While the current training offerings address some of the components found in the available malicious insider models, they cover only a fraction of what supervisors need to know to more accurately assess potential cyber insider threats. Second, supervisors have at least some knowledge of the daily behaviors of their employees. With appropriate training, they may pick up potential threat indicators that cannot be identified by technological means.

An equally important outcome of this work was to explore quantitative methods to measure engagement of the learner in a game-based learning environment, sometimes referred to as "immersive learning simulations" or "serious games". The eLearning Guild (2008) defines these solutions as "an optimized blend of simulation, game element, and pedagogy that leads to the learner being motivated by, and immersed into, the purpose and goals of a learning interaction. Serious games use meaningful contextualization, and optimized experience, to successfully integrate the engagement of well-designed games with serious learning goals." Game-based learning environments are believed to increase student engagement (Bergeron, 2005). However, measures of engagement are typically subjective. Psychophysiological measurements such as electroencephalography (EEG) and eye-tracking sensors have led to advances in scientists' understanding of the kind of learning retained. Through accelerated learning approaches and innovative ways to measure learner engagement, we hope to establish new guidelines for learning applications, just as one might use a user-in-the-loop iterative strategy when designing adaptive training systems (Fidopiastis & Nicholson, 2008).

This paper reports on two recent experiments conducted to investigate the design of TiER1's accelerated learning approach and improve our product. The first study compares the TiER1 accelerated training system (e.g., custom learning pathways, quizzes, game-based interaction, and feedback) to conventional multimedia training, while the second evaluates using objective psychophysiological measures to detect cognitive state changes of learners who receive the accelerated training system versus those who train using equivalent modules of conventional multimedia training.

STUDY I

The goals of Study I were: 1) to compare the XL-CITR training system to conventional multimedia instruction (CONVEN) using the same content; 2) to guide further improvements to the system's overall interface, content, and presentation; and 3) to inform the design of Study II. Data collection for the study focused on relative learning effectiveness, affective response to type of presentation, and time to completion of the instruction.

METHOD

Participants

With the assistance of the Colorado Homeland Defense Alliance (CHDA), a total of 33 participants in the Colorado Springs area were recruited to participate in this study. Colorado Springs was selected based on its proximity to several military installations and the cyber mission at Air Force Space Command. Throughout the recruitment process each participant was asked the same set of questions to gauge the specific criteria for which we were looking. For instance, participants were asked if they had previous supervisory experience or are currently supervisors as well as whether they have served or currently serve in the military. This was to ensure that each person in the study had an appropriate background for XL-CITR training.

Materials

The Program Quality Questionnaire (PQQ) was used to solicit the perceived quality of different aspects of the training program (e.g., interface design and interactivity) from participants.

Design and Procedure

Each of the 33 participants:

1) took a brief pre- and post-test of general cyber security knowledge,
2) completed one of two versions of a 2-4 hour training program on detecting insider threats to cyber security, and
3) participated in a group discussion. Participants were paid forty dollars as a way to compensate them for their time, which averaged approximately 180 minutes per participant.

Each qualified participant was randomly assigned to one of two groups (16 participants in the XL-CITR condition and 17 in the CONVEN condition). While the multimedia-based version (CONVEN) did not contain the role-play simulations provided by the XL-CITR system, the content of those simulations were replicated in web-based summaries, thus allowing the participants in the CONVEN group to experience the same material as those training with XL-CITR. At the beginning of the training session, each group completed a pre-assessment to provide individual baselines on the course's learning objectives. For the XL-CITR version, participants were accelerated into specific training lessons based on pre-established rules for performance on the pre-test. The CONVEN version simply captured the pre-assessment data and allowed participants to move through each course in the distance learning style normally administered during annual Air Force ancillary training. At the end of the training session, each group was given a post-assessment. Participants also completed the PQQ to assess their perceived quality of each training format.

Following the completion of the PQQ, all participants were brought together for a panel-facilitated discussion in which the purpose of the research and the experimental conditions were explained, and participants were encouraged to provide feedback on their respective training formats. The high-level points from the discussion were noted for later analysis and the feedback was used to improve the design of the XL-CITR version of the program before Study II.

Dependent measures for Study I included performance on both pre- and post-test assessments, gains in pre-to-post scores, responses to the assessment of program quality, and time to completion of the programs. The hypotheses for Study I were:

- H1: Post-test performance scores: XL-CITR Condition > CONVEN Condition.
- H2: Increases in pre-test to post-test performance scores: XL-CITR Condition > CONVEN Condition.
- H3: Participants' average ratings on the Program Quality Questionnaire: XL-CITR Condition > CONVEN Condition.
- H4: Average time to completion of program: XL-CITR Condition < CONVEN Condition.

RESULTS

Table 1 shows the demographics of the participants selected for the study. Participants were primarily male, over the age of 41, and had served in the military for over 15 years. All of the participants had experience supervising at some point in their career, but most were not currently supervisors.

Table 1. Demographic Data

Gender	Male- 81.8% (n=27)	
	Female- 18.2% (n=6)	
Age	19-25	6.2% (n=2)
	26-30	6.1% (n=2)
	30-40	6.1% (n=2)
	41-50	54.5 % (n=18)
	51-60	27.3% (n=9)
Years Served in the Military	0-5 years	6.1% (n=2)
	6-10 years	9.1% (n=3)
	16+ years	84.8% (n=28)
Currently Supervising	None	69.7% (n=23)
	1-5	18.2% (n=6)
	6-10	6.1% (n=2)
	11-15	3% (n=1)
	16+	3% (n=1)
Previously Supervised	1-5	15.2% (n=5)
	6-10	15.2% (n=5)
	16+	60.6% (n=20)

Table 2 presents the mean total time it took for participants to complete the XL-CITR and CONVEN version as well as mean pre- and post- test scores of each condition. Standard deviations are in parenthesis.

Table 2. Mean Time to Completion, Pre- and Post-Test Score for Both Training Programs

	Time to Completion	Pre-test score	Post-test score
XL-CITR Version	145.74 (53.33)	66.56 (11.12)	91.50 (7.01)
CONVEN Version	54.99 (9.96)	62.88 (10.32)	89.06 (6.10)

Hypothesis 1 (H1) was tested using an independent samples t-test ($t(31) = -1.07$, $p = .29$) which demonstrated that post-test performance scores were not significantly different across the XL-CITR version and CONVEN version.

The second hypothesis (H2) was examined using an independent samples t-test to evaluate the pre- to post-test performance scores across both conditions, and were not significantly different ($t(31) = .30$, $p = .77$).

The third hypothesis (H3) looked at the participants' ratings from the PQQ that was collected. The results revealed the mean values were slightly higher for the CONVEN condition on each of the dimensions but not significantly different than the XL-CITR condition.

Lastly, time to completion (H4) of the training program was significantly lower for the CONVEN condition (average completion time in minutes = 54.99, s.d. = 9.96) than for the XL-CITR condition (average completion time in minutes = 145.74, s.d. = 53.33).

DISCUSSION

The Study I results showed different findings than we expected. Both sets of participants increased their overall proficiency in the training content, with no significant variation in final scores or score increases between the two conditions. However, the participants in the XL-CITR condition took significantly longer than participants in the CONVEN condition. Because of the additional content and game segments in XL-CITR, it is not surprising that the CONVEN version was completed in a shorter time. Perhaps a more appropriate hypothesis should relate levels of learner engagement to the types of training material, instead of focusing on speed. Based on the length of the two different conditions (XL-CITR versus CONVEN), it was unrealistic to think that everyone would be experts and have accelerated out of enough modules to make it shorter than the CONVEN version.

For subjective results, CONVEN participants rated their reaction to the training presentation as slightly more positive than those participants in the XL-CITR condition. This result was initially puzzling until we evaluated the demographic data of participants. Over 81% of our participants in this first study were in the 41-60 age group. The more positive rating of the CONVEN condition may have been influenced by this age demographic. In panel discussions with participants, the older participants were not as interested in a game-based format. These older participants apparently preferred a training style that matched what they had historically been given within the Air Force. Throughout the panel discussions, many participants also highlighted that the younger generation in the military today would likely enjoy the game-based format in XL-CITR more than the older generation. In fact, of the six younger participants (19-40) they all seemed to enjoy the XL-CITR version more. These results heightened our awareness to a generational difference that exists surrounding media and training delivery methods. These results informed us that a diverse sample for Study II needed to be a priority for our own understanding of developing game-based learning environments in the Air Force and commercial world. As a result, our team moved the location of Study II, originally planned for the same Colorado Springs location, to Dayton, Ohio, where Wright-Patterson Air Force Base is located. Based on the diverse military and civilian personnel at Wright-Patterson, we assumed we would obtain a representative sample that would more closely match the intended population of this training.

STUDY II

A goal of augmented cognition research is to use psychophysiological measures such as EEG, pupil diameter and gaze tracking to identify, in real-time, perceptual, attentional, and cognitive workload states that may be helpful or detrimental to Warfighter performance (Schmorrow and Kruse, 2004). Recent investigations within military-relevant training environments suggest that EEG signatures of attention, memory and workload can be validly assessed during learning (Berka et al., 2007). Furthermore, EEG measures offer a reliable means to accurately quantify key aspects of information processing (Berka et al., 2004; Poythress et al., 2006). These studies suggest that changes in EEG power spectra, as well as event-related EEG changes, are identifiable and correlate with levels of skill acquisition in simple and complex tasks. Thus, these measures are useful in matching mental capabilities to task and learning requirements so that mastery over the subject matter is maximized while mental effort, when necessary, is minimized (Feldon, 2004).

Augmented cognition research specialists from the Interactive Simulation Laboratory at UAB helped perform the comparison study between the two refined training systems evaluated in Study I. The goal of this second study was to objectively assess the cognitive state of learners during the onset and the offset of specific design elements of XL-CITR training (fundamentals, quizzes, role-play simulations, and feedback) and compare those same quantitative measures with ones obtained from matching modules of a conventional multimedia version (implemented with e-learning software) using the same content. The result of such approach is an objective means of evaluating learning content and improving the effectiveness and efficiency of the iterative serious game development cycle.

An augmented cognition framework, using EEG and eye tracking sensors, was used to assess the workload and the engagement levels of learners as they progressed through abbreviated, generalized versions of XL-CITR and the CONVEN version used in the previous study. This assessment allowed design elements to be mapped onto intra- and inter-participant psychophysiological differences. That is, the experiment established individual baselines for the psychophysiological response measures and identified changes in them during randomly assigned onset and offset exposure to the various program design elements. If successful, the pattern of these changes in workload and engagement can be used to guide subsequent iterations of XL-CITR development.

METHOD

Participants

A total of 31 male and female volunteers from the local military population (active duty, retired, or government-civilian) in Dayton, OH participated in this study (demographic data shown in Table 3). Participants were recruited through the local base networks and business groups.

Table 3. Demographic Data

Gender	Male- 80% (n=24)	
	Female- 20% (n=6)	
Age	19-25	16.7% (n=5)
	26-30	20% (n=6)
	30-40	26.7% (n=8)
	41-50	23.3 % (n=7)
	51-60	6.7% (n=2)
	61+	6.7% (n=2)
Years Served in the Military	0-5 years	43.3% (n=13)
	6-10 years	16.7% (n=5)
	11-15 years	6.7% (n=2)
	16+ years	20% (n=6)
	None	13.3% (n=4)
Currently Supervising	None	77% (n=23)
	1-5	16.4% (n=5)
	6-10	3.3% (n=1)
	11-15	3.3% (n=1)
Previously Supervised	1-5	16.7% (n=5)
	6-10	13.3% (n=4)
	11-15	6.7% (n=2)
	16+	26.7% (n=8)
	None	26.7% (n=8)

As mentioned previously, the demographics and PQQ results of Study I informed the location change to Dayton, Ohio to ensure a more diverse sample. Based on the results, the age range was well-represented over the four conditions.

Materials

Subjective Self-Assessments: The same survey of program quality (i.e., PQQ) evaluated in Study I was used in this study. In addition, the NASA- TLX, a subjective workload assessment using a multi-dimensional rating procedure, was added. The NASA-TLX derives an overall workload score based on a weighted average of ratings on six subscales, including Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort and Frustration. In this study, we were interested in the subjective scores of subscale ratings. Participants responded twice, once for the XL-CITR module and once for the conventional multimedia training module. These subjective measures supported the psychophysiological measures of workload.

Equipment

EEG. The B-Alert X10 wireless EEG system developed by Advanced Brain Monitoring (ABM) was used in this study. The system combines a battery-powered headset with a sensor placement system, following international standards. The full system is a lightweight, easy-to-apply cap that can acquire and analyze ten channels of high-quality EEG data. Monopolar sensor site locations on the system included: F3, F4, C3, C4, P3, P4, Fz, Cz, Poz, with mastoid references. The head pack contained miniaturized electronics that amplify, digitize, and wirelessly transmitted the EEG data.

Eye Tracking. Blink rate, pupil dilation, and scan path are convergent measures for mental workload (Sciarini, Fidopiastis, and Nicholson, 2009). The Arrington Research Viewpoint binocular eye tracker system was used to track eye movements during the study. The eye tracker system uses a wearable eyeframe with eyecamera and the illuminator integrated into it as shown in Figure 1a. It allows the system to dynamically track as the operator changes head position.

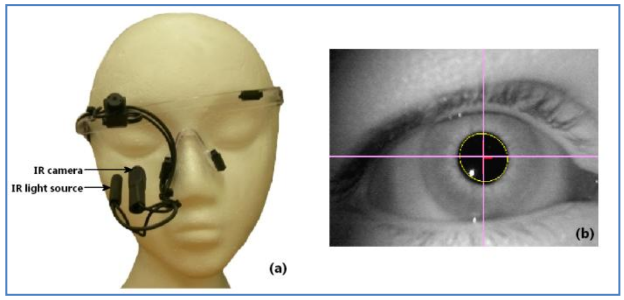

Figure 1. (a) Arrington Research Viewpoint eye tracker system. (b) Pupil center detection for accurate eye tracking.

Design and Procedure

Upon entering the testing room, participants completed all IRB approved documents. Participants were then fitted with the wireless EEG. An impedance check of all sensors confirmed that the sensors settled to under 40 Kohm. Participants then performed a set of baseline tasks (a 3-choice vigilance task, an eyes-open rest task, and an eyes-closed task) to calculate the cognitive states of engagement and workload of the learner. Once the EEG baseline was completed, the participant underwent a calibration procedure to align the eye tracker coordinate system to the monitor displaying the XL-CITR content. In the calibration procedure, participants directed their gaze to 16 predefined points, one at a time, on the display, while the eye tracker recorded their eye position.

As in Study I, two versions of the same training content were used and displayed in different orders depending on the condition. Participants were exposed to both a lesson from the XL-CITR training (XL-CITR Module A or B) and a lesson from the conventional multimedia training (CONVEN Module A or B). Each person was randomly assigned to one of four experimental conditions (GRP 1-4), balancing order of presentation and content set:

- SII-GRP1: XL-CITR Module A, then CONVEN Module B
- SII-GRP2: XL-CITR Module B, then CONVEN Module A
- SII-GRP3: CONVEN Module A, then XL-CITR Module B
- SII-GRP4: CONVEN Module B, then XL-CITR Module A

Dependent Measures

The dependent measures for this study included the pre- and post- tests, the EEG measures (EEG percent workload, EEG percent engagement, frequency and intensity changes at each EEG location), the eye tracking measures (blink rate, pupil diameter), the subjective assessment of the PQQ, and the perceived workload based on the NASA-TLX subscales.

RESULTS

Testable Assumptions (TA) and Hypotheses (H)

SII-TA1: The overall EEG measure of percent engagement will provide a measure of comparative differences between XL-CITR and CONVEN.

SII-TA2: The overall EEG measure of workload will assist in further differentiating between performance on XL-CITR and CONVEN.

SII-TA3: The NASA-TLX perceived magnitude of load subscales (mental, physical, temporal, own performance, effort and frustration) will provide convergent evidence of effects found by the EEG measures.

SII-H1: Average scores on EEG measures of percent engagement: XL-CITR > CONVEN.

SII-H2: Average scores on EEG measures of workload: XL-CITR < CONVEN.

SII-H3: Perceived magnitude of load for subscales mental, physical, temporal, effort and frustration: XL-CITR < CONVEN. Own performance: XL-CITR > CONVEN.

SII-H4: Participants average ratings on the Program Quality Questionnaire: XL-CITR Condition > CONVEN Condition.

Prior to training with XL-CITR or CONVEN, participants scored similarly on the pre-test, average 64, s.d.= 23.84 and 63, s.d. = 13.34, respectively. After training on XL-CITR first participants scored an average 5 points higher on the post-test, 91, s.d. = 7.81 versus 87, s.d. = 11.14. However, this result was not a statistically significant difference.

Table 4 shows the means and the standard deviations for questions on the PQQ that mark significant differences in participants' perceptions about each type of training method. Each 5 point Likert scale question designates positive responses as 5 and 1 as most negative. A one-way ANOVA was performed to identify those questions indicating significant differences between the two types of training. The results of the ANOVA are also represented in Table 4. These results show partial support for SII-H4.

Table 4. Program Quality Questionnaire

Program Quality Questionnaire	Training	
	XL-CITR	**CONVEN**
Interactivity: Rate the users interaction with training	4.23(1.01)	3.23(1.28)
	F(1)= 11.34 p = .01	
I would choose this format in the future for learning about cyber security.	4.1 (.92)	3.5 (1.25)
	F(1) = 4.46 p = .04	
How would you rate the level of interactivity of the training?	4.7 (.70)	3.9 (1.05)
	F(1) = 12.06 p = .01	
The training held my attention.	4.03 (.61)	3.6 (1.00)
	F(1) = 4.07 p = .05	
I found the training interesting.	4.43 (.68)	3.9 (.84)
	F(1) = 7.26 p = .01	

For questions rating training methods on interactivity, attention, and interest, participants reported significantly higher scores for the XL-CITR training. Further, 66% of the participants who experienced the XL-CITR Module A reported that they would choose this format in the future for learning about cyber security, as opposed to 31% who recommended the CONVEN Module B. Fifty-five percent of those who experienced XL-CITR Module B reported that they would choose this format for future learning, while 49% would recommend CONVEN Module A.

The average scores for the perceived magnitude of load for the NASA-TLX subscores, as well as the EEG measures of percent engagement and workload are displayed in Table 5. These results were further analyzed using a two-way 2 (Training types: XL-CITR, CONVEN) x 4 (Training Version) mixed ANOVA with repeated measures on the training type variable. There were no significant main effects or interactions for physical, temporal, and frustration subscores of the NASA-TLX. As well, there was no effect of EEG engagement.

Table 5. Mean Scores for NASA-TLX and EEG Metrics

NASA-TLX	**XL-CITR**	**CONVEN**
Mental	54.00 (19.80)	39.50 (23.06)
Physical	20.83 (14.09)	19.17 (13.27)
Temporal	31.17 (20.03)	28.17 (17.49)
Performance	30.67 (20.37)	27.50 (15.74)
Effort	50.83 (23.20)	38.67 (23.08)
Frustration	32.17 (21.96)	28.17 (19.36)
EEG	**XL-CITR**	**CONVEN**
High Engagement	46.7093 (11.60)	45.7093 (11.55)
Workload	51.063 (8.86)	49.50 (8.747)

There was a main effect of Training Type on mental demand, $F(1, 26)=13.07$, $p = .001$ and effort, $F(1, 26) = 6.30$, $p = .019$. Participants perceived their mental demand and effort greater when training using XL-CITR. There was also an interaction between Training Type and version of the training on the performance subscale, $F(3, 26) = 4.40$, $p = .012$. Performance on XL-CITR was perceived as better than CONVEN for Modules A and B of the testing conditions. The EEG metric of workload, showed a main effect of Training Type, $F(1, 26) = 8.81$, $p = .006$, where participants training with XL-CITR showed overall higher workload.

DISCUSSION

While results from the pre- and post-test scores did not show significant differences between XL-CITR and CONVEN, participants reported that they prefer XL-CITR as a future delivery system for cyber threat detection training. Participants also reported that they found the XL-CITR platform more interactive, interesting, and captivating than CONVEN. The EEG measure of engagement monitored the changing demands for sensory processing and attention resources during each training session. Despite participant reports of being more engaged in the XL-CITR training, the metric did not differentiate between the two training types. Thus, SII-H1 was not supported by this study.

The results of the NASA-TLX mental demand and effort suggest that persons who trained using XL-CITR perceived their mental and perceptual activity as higher and that they had to mentally work to accomplish their level of performance. This higher mental workload was objectively supported by the EEG workload measure. While EEG workload does distinguish between XL-CITR and CONVEN and provides support for SII-H2, it does so in the opposite direction, as do the NASA-TLX subscale measures of mental demand and effort. The NASA-TLX performance subscale does partially support that persons perceive their performance as better than traditional training supporting SII-H3, but only under certain conditions.

There were three important outcomes of this study:
1) sensors such as the EEG can be used in field training exercises and evaluation with minimal intrusion into the training environment; 2) psychophysiological measures such as EEG can provide objective measures of overall, as well as specific comparisons between training types; and 3) underlying theories used to guide the use of psychophysiological measures for training evaluation (e.g., Cognitive Load Theory-CLT) may need further refinement as more field studies and new results from the field of Neuoroeducation provide better information on how the brain learns. For example, while CLT would suggest that efficient learning is indicative of lower workload (Kalyuga, Chandler, and Sweller, 1999), our results suggest that mental demand and effort may play a role in positive training effects during single training efforts. In this regard, arousal or a state of alertness may not be indicated solely by measures of attentional engagement (e.g., EEG), but by the other measures such as perceived mental effort of the learner. This research does show that objective measures such as EEG workload can provide support for qualitative data as well, thus providing convergent evidence as to the actual cognitive state of the learner.

On all PQQ questions of interactivity, participants rate XL-CITR as better than CONVEN. Cognitive Theory of Multimedia Learning (CTML) suggests that interactivity is a key component to deeper learning of multimedia-based training (Mayer and Moreno, 2003). The level and type of interactivity, may lead to higher positive cognitive workloads that can become objective metrics when evaluating serious game design. This aspect of game development should be explored in future work.

Further, Kahneman (1973) suggests that the more mental effort we expend on a task, the more attention we place on that activity. This heightened attention necessary for improved performance may not be totally explicated by a single psychophysiological measure. Our analysis of the eye tracking data (e.g., pupil diameter changes and gaze pattern) may lend convergent support that arousal as defined as mental effort is a valid metric for evaluating serious games for training. Taken together with the EEG metrics, we may be able to develop a methodology for best practices in evaluating serious game-based trainers for military applications using psychophysiological measures.

Based on the results of Study II, next steps in evaluating the XL-CITR platform include: 1) understanding what aspects of interactivity lead to better transfer and retention of the training material; 2) adding an assessment for transfer and retention of training; and 3) developing guidelines for serious game development that requires in field testing of software.

ACKNOWLEDGEMENTS

This material is based upon work supported by the Air Force Research Laboratory (AFRL) under Contract No. FA8650-10-C-6061. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of AFRL.

REFERENCES

Bergeron, B. (2005). *Developing serious games*. Hingham, MA: Charles River Media.

Berka, C., Levendowski, D., Lumicao, M., Yau, A., Davis, G., & Zivkovic, V. (2007). EEG correlates of task engagement and mental workload in vigilance, learning and memory tasks. *Aviation Space and Environmental Medicine, 78(5)*, B231-B244.

Berka, C., Levendowski, D. J., Cvetinovic, M., Petrovic, M. M., Davis, G. F., Lumicao, M. N., Popovic, M. V., Zivkovic, V. T., Olmstead, R. E., & Westbrook, P. (2004). Real-time analysis of EEG indices of alertness, cognition and memory acquired with a wireless EEG headset. Special Issue of the *International Journal of Human-Computer Interaction on Augmented Cognition,17(2)*,151-170.

Department of Defense (2009). Department of Defense fiscal year (FY) 2010 budget estimate. Retrieved from: http://www.darpa.mil/Doc/2010 PBDARPA May2009.pdf

eLearning Guild (2008). Immersive Learning Simulations. Retrieved from http://www.elearningguild.com/ research/archives/index.cfm?action=viewonly2 &id=128

Feldon, D. F. (2004). The benefits of efficiency metrics in the evaluation of simulation task performance. Paper presented at the *Annual Meeting of the American Education Research Association*. San Diego, CA.

Fidopiastis, C. & Nicholson, D. (2008). User-in-the-loop adaptive system design: Information fusion examples for visualizing and measuring cognitive states. *Applied Ergonomics International Conference*. Las Vegas, NV.

Identity Theft Resource Center (2011) Retrieved from: http://www.idtheftcenter.org/artman2/publish/lib_survey/ITRC_2008_Breach_List.shtml

Jacobson, S. (March-April, 2005). Can we control training waster? *Frontline Magazine*, 30-31.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, New Jersey: Prentice-Hall.

Kalyuga, S., Chandler, P. & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351-371.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*, 43–52

Moe, T. & Chubb, J. (2009). Liberating learning: Technology, politics and the future of American education. Hoboken, NJ: Wiley (Jossey-Bass).

Office of Management and the Budget (2005). Report to Congress on the Implementation of the E-Government Act of 2002. Retrieved from: http://www.whitehouse.gov/OMB/inforeg/2004_egov_report.pdf

Poythress, M., Russell, C., Siegel, S., Tremoulet, P., Craven, P., Berka, C., Levendowski, D., Chang, D., Baskin, A., Champney, R., & Hale, K. (2006). Correlation between expected workload and EEG indices of cognitive workload and task engagement. In D. Schmorrow, K. Stanney & L. Reeves (Eds.), *Augmented cognition: Past, present and future* (pp. 32-44). Arlington, VA: Strategic Analysis, Inc.

Schmorrow, D. D. & Kruse, A. A. (2004). Augmented cognition. In W.S. Bainbridge (Ed.), *Berkshire encyclopedia of human-computer interaction* (pp. 54-59). Great Barrington, MA: Berkshire Publishing Group.

Verizon (2011) 2011 Data breech investigations report. Retrieved from: http://www.verizonbusiness. com/resouces/executivesummary/es_2011-data-breach-investigations-report_en_xg.pdf.