2011

CRAM It! A Comparison of Virtual, Live-Action and Written Training Systems for Preparing Personnel to Work in Hazardous Environments

Catherine Stocker
University of Pennsylvania, cstocker@seas.upenn.edu

Ben Sunshine-Hill
University of Pennsylvania, bsunshin@seas.upenn.edu

John Drake
University of Pennsylvania, drake@seas.upenn.edu

Ian Perera
University of Pennsylvania

Joseph T. Kider
University of Pennsylvania, kiderj@seas.upenn.edu

See next page for additional authors

Recommended Citation

Stocker, C., Sunshine-Hill, B., Drake, J., Perera, I., Kider, J. T., & Badler, N. I. (2011). CRAM It! A Comparison of Virtual, Live-Action and Written Training Systems for Preparing Personnel to Work in Hazardous Environments. *IEEE Virtual Reality Conference (VR)*, 95-102. http://dx.doi.org/10.1109/VR.2011.5759444

The 2011 IEEE Virtual Reality Conference (VR) took place March 19-23, 2011 in Singapore.

CRAM It! A Comparison of Virtual, Live-Action and Written Training Systems for Preparing Personnel to Work in Hazardous Environments

Abstract
In this paper we investigate the utility of an interactive, desktopbased virtual reality (VR) system for training personnel in hazardous working environments. Employing a novel software model, CRAM (Course Resource with Active Materials), we asked participants to learn a specific aircraft maintenance task. The evaluation sought to identify the type of familiarization training that would be most useful prior to hands on training, as well as after, as skill maintenance. We found that participants develop an increased awareness of hazards when training with stimulating technology – in particular (1) interactive, virtual simulations and (2) videos of an instructor demonstrating a task – versus simply studying (3) a set of written instructions. The results also indicate participants desire to train with these technologies over the standard written instructions. Finally, demographic data collected during the evaluation elucidates future directions for VR systems to develop a more robust and stimulating hazard training environment.

Keywords
virtual reality, training, interactive environments, evaluation/methodology

Disciplines
Computer Sciences | Engineering | Graphics and Human Computer Interfaces

Comments
The 2011 IEEE Virtual Reality Conference (VR) took place March 19-23, 2011 in Singapore.

Author(s)
Catherine Stocker, Ben Sunshine-Hill, John Drake, Ian Perera, Joseph T. Kider, and Norman I. Badler

CRAM it! A Comparison of Virtual, Live-Action and Written Training Systems for Preparing Personnel to Work in Hazardous Environments

Catherine Stocker Ben Sunshine-Hill John Drake Ian Perera Joseph T. Kider Jr. Norman I. Badler

University of Pennsylvania

ABSTRACT

In this paper we investigate the utility of an interactive, desktop-based virtual reality (VR) system for training personnel in hazardous working environments. Employing a novel software model, CRAM (Course Resource with Active Materials), we asked participants to learn a specific aircraft maintenance task. The evaluation sought to identify the type of familiarization training that would be most useful prior to hands on training, as well as after, as skill maintenance. We found that participants develop an increased awareness of hazards when training with stimulating technology – in particular (1) interactive, virtual simulations and (2) videos of an instructor demonstrating a task – versus simply studying (3) a set of written instructions. The results also indicate participants desire to train with these technologies over the standard written instructions. Finally, demographic data collected during the evaluation elucidates future directions for VR systems to develop a more robust and stimulating hazard training environment.

KEYWORDS: Virtual Reality, Training, Interactive Environments, Evaluation/Methodology

INDEX TERMS: I.3.6 [Computer Graphics]: Methodology and Techniques--Interaction Techniques; K.3.1 [Computing Milieux]: Computers and Education – Computer Uses in Education;

1 INTRODUCTION

Occupations that involve interaction with large machinery, such as building construction, auto repair or aircraft maintenance, present many hazardous situations and make correct task performance fundamental to trainee safety and equipment integrity. In these environments, proper initial skill acquisition is a crucial but dangerous procedure. The ability for novices to train virtually, prior to hands-on interaction, could prevent injuries to personnel as well as excess wear and expensive damage to equipment – all without the oversight of a live human trainer.

In our evaluation, we investigate which type of training can make personnel most aware of the hazards in a work environment prior to live interaction with the danger. We compared three different learning methods and discovered both the video- and simulation-based methods produced significantly better hazard awareness than the traditionally-used written descriptions required in our selected training application. In addition, when surveyed for preferences, participants also indicated a desire to *train* with these technologies over written instructions, even though they understood that explicitly following the written instructions is mandated by the application.

e-mail: {cstocker, bsunshin, drake, eperera, kiderj, badler} @seas.upenn.edu

The software model we used for training was designed in conjunction with the United States Air Force. The content of the system was based on a standard textual Technical Order (TO) used by Air Force maintenance students during their training to become flight maintainers. Our findings should be applicable, however, to all virtual training domains that require increased hazard awareness during interaction with physical tools and machinery.

Section 2 provides background for the context of the research effort. Section 3 describes the VR training system we developed. Section 4 explains the methods, assumptions and procedures of the evaluation. Section 5 offers the results and discussion, and Section 6 provides conclusions from the research while proposing recommendations for future work. The complete evaluation plan, along with consent forms and questionnaires can be found on the research webpage: http://cg.cis.upenn.edu/hms/research/CRAM

2 BACKGROUND

There exist a number of relevant attempts at training systems using virtual and augmented reality [1, 2, 3, 4] as well as games that train in military procedures [5, 6]. These have been well surveyed in previous literature [7]. However, experimental evidence has been ambiguous about the advantages of technology for improving training and performance. Studies show that in *certain* situations, simulation-based training yields better results than conventional learning methods [8, 9], but in other situations, simulations must be augmented with real-world instruction to improve training effectiveness [10]. In addition, simulations may be more appropriate for certain populations [11, 12], and are best used in learning environments that aim to teach *intuitive* understanding rather than rote memorization [13, 14].

It has been demonstrated that students with a predisposition for, or incentives to, learning are more likely to seek out educational opportunities, persist in the face of difficulty, have greater cognitive flexibility and retain material longer [15, 16, 17]. Therefore we believe a training aid system needs to *attract* participation so that the system is actually used. Our goal was to answer two questions. Will using a VR training system (such as the one we created) cause users to become more aware of the hazards involved in a maintenance task, or are simpler methods sufficient (such as reading the TO text or watching a video of the instructor describe and demonstrate the system)? Which method would users *prefer* to practice with?

There are various ways to test the knowledge acquired in training simulations – each specific to the type of knowledge trying to be instilled. If the goal is to evaluate memorization, evaluators may track performance time and number of errors along repeated trials in the given virtual task [18]. If the newly-acquired skill can only be judged visually, the test may involve expert analysis of the learned movements [19]. The focus of our evaluation was not to have the participant memorize the steps of a procedure or specific movements, but to understand the reasoning behind the steps as they relate to the functionality of the aircraft and to gain a thorough understanding of the hazards involved in order to prevent damage to aircraft or injury to personnel. For that reason, the knowledge acquisition of our simulation was tested using our own unique method: showing a set of video clips to the trainees and asking them to indicate any hazards they observed with an explanation of why they are dangerous.

There are also numerous ways to test the usability of a system [20, 21, 22, 23], each with its own advantages and disadvantages [24]. We chose to administer the System Usability Scale (SUS), a 10-item Likert scale questionnaire intended to provide a quick and easy method to obtain user satisfaction with a variety of software systems [25]. The scale is a popular measure of the usability of virtual education environments [19, 26] and has been shown to be one of the more reliable usability questionnaires when used with small sample sizes [27]. The SUS questions were added to a subjective questionnaire we designed to elucidate preferences for the three training aids (interactive simulation, videos, TO text) we were comparing, which uses a combination of Likert scale questions, essay questions and ordering questions.

3 Overview of CRAM

We designed and implemented an interactive software system, called CRAM (Course Resource with Active Materials) that was used in our evaluation. CRAM was built with the intention of:

- Capturing the instructional experience of a course instructor in a computer (software) based training environment.
- Providing a mechanism for storing, archiving and efficiently accessing course materials, simulations, TO's, multimedia materials and instructional expertise.
- Allowing several trainees to cooperate in a simulated training task, in order to more closely approximate real-world team-based roles.
- Utilizing human model avatars as coaches, communicators (instructor surrogates) or simulated maintainers to instruct and illustrate correct and incorrect procedures and practices - especially cautions and warnings.

CRAM is implemented as a web-based client/server application. A centralized server stores all training content, and clients with the appropriate multimedia plugins connect and choose the instructional unit to train. The server additionally serves as a matchmaking and coordination hub for trainees participating in multi-user simulations, although much of the communication during the simulation is peer-to-peer.

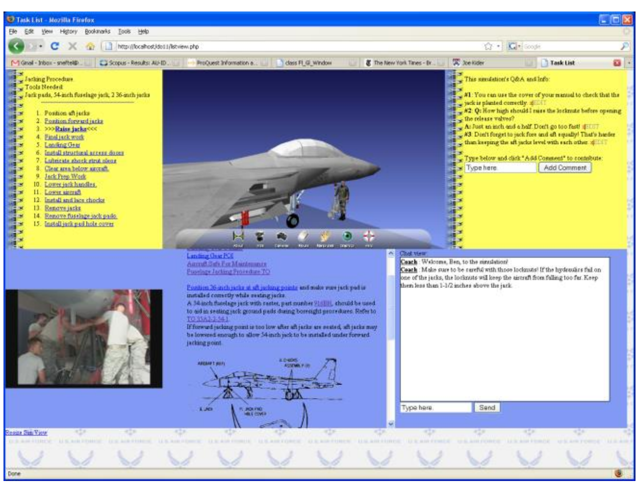

Fig. 1 The CRAM trainee interface. Clockwise from upper left: The task list panel, the interactivity panel (3D models), the wiki panel, the chat panel, the step details panel, and the video panel.

3.1 Trainee Interface

The interface through which trainees interact with the CRAM system is shown in Fig. 1. The design of the interface is oriented around the concept of giving a trainee various tools to complete a given procedure in the interactivity panel. The task list panel located on the left side of the interface displays the top-level sequence of steps to be performed. At any time, the user can use the task list panel to monitor his progress through the procedure. The next step, which must be completed to advance through the procedure, is clearly indicated. Additionally, all steps are clickable; the user can click a step to view its details in the step detail panel, but by default that panel displays the details for the next step to be executed. These details are taken directly from the official TO text; they may be supplemented, however, with reminders and advice from the instructor. These details are fully hyperlinkable, and can display multimedia content, such as videos or hazard simulations, in the video panel to the left. A chat panel is used to communicate with other trainees and the virtual coach. Chatting provides easy textual communication during multi-user simulations, and a wiki panel allows long-term retention of informal information about the procedure. Finally, the interactivity panel allows full 3D manipulation of the equipment involved in the training procedure.

For instance, a trainee may begin the procedure by clicking through the steps in order to see what needs to be done. In the study, the task was to safely jack an airplane, so the first steps include 1) positioning the aft jacks 2) positioning the forward jacks and 3) raising the jacks. Each of these steps have a number of specific instructions and warnings, which can be viewed in the step detail panel. A warning may indicate injury to the trainee or the plane. For example, the system warns that positioning the jacks out of order (forward before aft) could cause the front of the plane to be punctured because the plane may settle onto the forward jack. The steps also may have associated video, such as a demo of the jacks being placed, so that the trainee could watch before beginning. After doing so, he performs the step in the interactive pane by using the keyboard to move his virtual character appropriately. The system, recognizing that a step has been completed properly, advances the simulation and displays the details for the next step. The trainee can go on to complete that step, but may first review the details of the first step if desired**.**

3.2 Content Representation

A *procedure* in CRAM involves the multimedia content of the simulation, such as meshes and textures for the objects being manipulated, as well as textural information on the procedure itself and how it is carried out. This content is separated into three layers which intercommunicate to provide high-level behavior and allow easy content creation and modification without extensive rewriting.

1) The *physical layer* consists of the actual 3D content, as well as simple information about semantically meaningful sites and objects. For instance, the physical layer for the aircraft jacking procedure task used in the evaluation includes a test to determine whether the jack ram is below the aircraft jack pad at any given moment.
2) The *practical layer* uses the information from the physical layer to maintain the state of the world. Unlike the physical layer, this does not merely include the physical configuration of the objects in space, but domain-specific information on those objects. For instance, the practical layer for the aircraft jacking procedure determines whether, dur-

ing jacking, the aircraft has become imbalanced to the extent that it is in danger of falling off the jacks.
3) The *procedure layer* uses information from the practical layer to track the trainee's progress through the training procedure. This layer determines when the trainee has proceeded to a new step in the procedure, and can raise warnings when the trainee performs steps in an incorrect order or in an order that will lead to a dangerous situation.

The physical layer and practical layer do not explicitly refer to the procedure being trained, and may be reused fully for several related procedures. The three layers are arranged in decreasing order of the technical acumen required to create them: Physical layer content requires the use of standard 3D modeling tools, whereas the practical and procedure layer content can be authored with a domain-specific tool with minimal technical knowledge of the underlying CRAM architecture.

4 Evaluation

Aircraft maintenance is a core function performed by numerous personnel in the United States Air Force. This function encompasses activities such as flight line servicing, equipment repair, and the training of new recruits. Our evaluation scenario focused on familiarization training for one aircraft maintenance task: the F-15 aircraft jacking procedure. Aircraft jacking is an inherently hazardous task with multiple critical steps, requiring several individuals working in a team for safe execution and is essential to many subsequent maintenance tasks.

The procedure involves placing three jacks under specific aircraft jacking pads, then raising the jacks in a manner so as to keep the aircraft level at all times. Hazards abound: if the nose jack is put in position before the aft jacks are correctly seated, the aircraft can rock forwards and be damaged by the head of the nose jack, and if the jacks are not properly level on the ground, the aircraft can rock to one side as weight is taken off the wheels, leading to a collapse. During the jacking process, it is necessary to regularly tighten a locknut on each jack, so that in the event of hydraulic failure of one of the jacks, the aircraft would not drop more than a fraction of an inch. These and other hazards are latent in the procedure to be performed. Each one becomes a potential danger if instructions are not correctly followed, and the potential dangers of deviating from the instructions in a particular way are not obvious without a detailed understanding of the systems in question.

The user interacts with the aircraft jacking simulation in CRAM through detailed three-dimensional models of an aircraft and jacks. The same operations with the models can be performed, though they are not engaged in the same way (e.g. a keyboard stroke instead of a physical turning of a locknut). Small visual cues, such as the jack pads turning a color when correctly positioned, are used to compensate for the lack of depth perception that is normally important to such procedures. The system allows an unlimited number of trainees to view or participate in the training task simultaneously.

The standard training procedure for the F-15 aircraft jacking procedure involves classroom instruction followed by the reading of the TO prior to and during hands-on-training. The TO text specifies the task step-by-step and includes annotated drawings of equipment and warnings, cautions, and hazards. It is our hypothesis that trainees of a 3D, team-based, technical task such as this could become better informed of the hazards, cautions and warnings while training in a virtual environment over reading a TO or watching a video of an instructor performing the task. We also hypothesize that they would prefer this VR training method.

To test these hypotheses we split participants into 3 training groups: CRAM, video, and TO. After participants trained with their assigned method, they filled out both a quantitative questionnaire and subjective questionnaire. The quantitative questionnaire required them to identify hazards in a scene, while the subjective questionnaire asked them to specify their preferred method of training.

4.1 Participants

A total of 48 Air Force technical training students, assigned to the 82nd Training Wing at Sheppard Air Force Base and enrolled the F-15 Fighter Aircraft Maintenance Apprentice Course (J3AQR2A333A025A), participated in the study. The group was composed of six different classes, spanning three "blocks" (levels) of training. Students from these 3 blocks (numbers 3-5) were chosen such that half of the participants had already learned the task we were training (i.e., they were refreshing their skills; we labeled them the "experienced" or "have jacked" group) and half had not yet learned it (the "novice" or "have-not jacked" group). The average age of the participants was 19.9 (range 18-25) and all were male. The assignment of participants to the CRAM, video or TO groups was random.

Participation did not entail any compensation – classes of students volunteered in exchange for not attending class, and were made aware that their involvement was voluntary and terminating their session would not result in prejudice to present or future care or services at the university or within the United States Air Force.

The data from one participant was discarded after he discovered a small bug in the CRAM software that significantly delayed his progress. No other participants triggered this bug and thus it did not affect any other participant performance data.

4.2 Materials

Equipment use varied by training group. Members of the CRAM condition trained on standard consumer-grade Windows-based laptops provided by the investigators, preconfigured with the CRAM software. Participants assigned to the video condition viewed a video on a large television screen as a group but were not allowed to talk during the showing. In the TO condition, students reviewed a hardcopy version of the aircraft jacking procedure training manual.

All participants viewed a final video as a group on a large television screen. The video contained 9 clips of F-15 aircraft jacking procedures that may have involved hazards. To ensure that no group had an advantage in demonstrating their intuitive knowledge based on the medium that they trained on, the 9 clips were divided into 3 media types:

1) Text description of a situation such as:

   After seating the nose and aft jacks (assume they were seated correctly), maintenance personnel raise each jack up (assume all jacking precautions are taken), one by one, in the order they were seated.

2) Video clips of steps being performed, such as a clip that may have been cut up to give the impression that the maintainer is doing something dangerous, or left whole, showing a safe step completion.

3) Animations that have a virtual agent performing a step using motions we acquired by motion capturing an expert (a Staff Sergeant at Sheppard AFB who teaches the F-15 jacking procedure) performing the task. These motions have either been left intact to show the agent demonstrating the step correctly and safely, or have been altered to show the agent performing a step dangerously.

Each media type was represented in three scenes of the objective test video. Two of the three appearances demonstrated hazards and one showed no hazard.

4.3 Procedure

Class sizes participating in the evaluation ranged from seven to nine students. Each class was randomly split-up upon arrival into three training groups of approximately the same size (i.e. two to four students per group depending on the class size). Prior to the start of the evaluation, all students completed a standard consent form and a demographic questionnaire. After completing the demographic questionnaire, all students were given an instruction sheet describing their task and 35 minutes to complete the task.

Participants in the CRAM condition used CRAM to step through the task in a virtual environment. Their instruction sheets provided brief explanations of how to successfully interact with CRAM and explained the multi-media content they could explore. None had any prior experience with the CRAM software.

The video condition group was instructed to watch a video of an actual instructor describing and demonstrating the F-15 jacking procedure in a hangar to a class. Some asked to take notes, which they were allowed to do, but they could not use their notes during the testing phase.

Those in the TO condition read the TO description of the F-15 jacking procedure. The provided TO material is used by all current Air Force maintenance students during their training as study material before and during hands-on-training.

In order to produce the most realistic results, all participants were asked to study as if they were preparing to go out to the hangar for hands-on training at the completion of the experiment.

After learning the procedure, the three groups were assembled together and shown the evaluation video. After each segment was shown, students were asked to document in their objective questionnaire what, if any, hazard they perceived to be present.

In order to confirm the validity of the expected answers, two experts (maintenance instructors) were also asked to take the objective test. The expert responses to questions 5, 6 and 7 did not match the expected answers and it was decided that because those questions presented ambiguous scenarios, they should be withheld from the grading of the evaluation. Therefore, only questions 1-4 and 8-9 were considered in the grading of the students' scores and the total possible number of correct answers was 6. Because the excluded questions were each referencing a different media type (animation, text and live action respectively), no group gained an unfair advantage by their exclusion.

After finishing the objective test, participants filled out a subjective questionnaire tailored specifically for each of their respective groups, which contained qualitative Likert questions such as (for a CRAM user):

Practicing a task in CRAM could help me become more aware of the hazards involved in a maintenance task.

Besides these, there were essay questions to supplement our understanding of what the participant experienced and a request for the student to rank which method of training they would turn to first if they needed to improve their proficiency with a task. In addition, the CRAM condition had 10 extra questions about the usability of the system, taken from the System Usability Scale questionnaire [25].

After completing the subjective questionnaire, participants were debriefed and released.

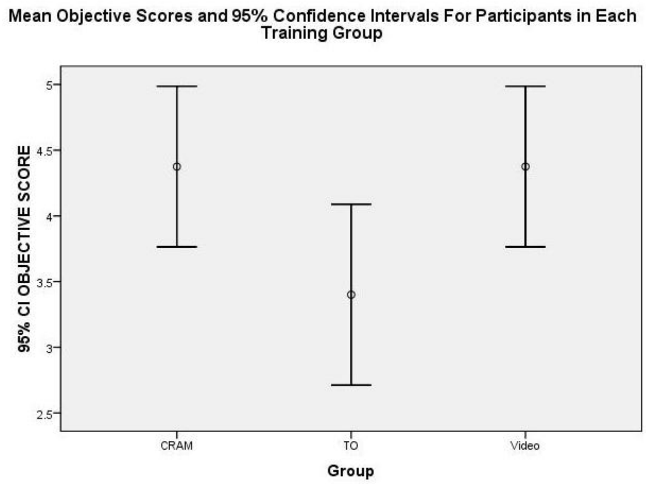

Fig. 2. Mean objective scores and 95% confidence intervals for Participants in each training group

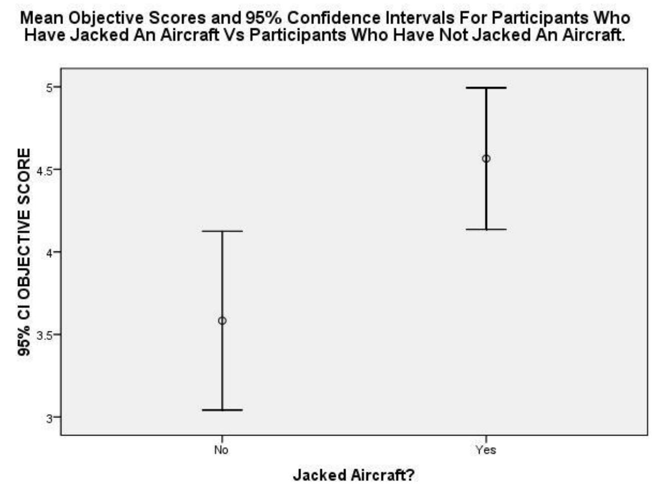

Fig. 3. Mean objective scores and 95% confidence intervals for novice ("no") versus experienced ("yes") participants

5 Results and Discussion

5.1 Objective Scores

A one-way analysis of variance (ANOVA) was calculated on the scores of the participants from the objective test, with the independent variable of assigned training method. As Fig. 2 illustrates, the effect of training method was significant ($F(2,44)=3.50$, $p=0.039$). Participants who trained on either CRAM or the video ($M=4.38$, $SD=1.15$ for both groups) scored higher on the objective test than participants who trained by reading the TO ($M=3.40$, $SD=1.24$).

A number of additional factors were examined to determine their influence on participants' objective test score. Three of those variables proved significant. First, a one-way ANOVA showed that the effect of previous jacking experience on a participant's objective test score was significant (F(1,47)=8.566, p=0.005) (Fig. 3). Participants who had previous jacking experience (M=4.57, SD=0.992) scored higher on the objective test than participants who had no previous jacking experience (M=3.58, SD=1.283). In addition, current block of study produced a significant influence on participants' objective test score (F(2,44)=6.118, p=0.005), which could be related to the differ-

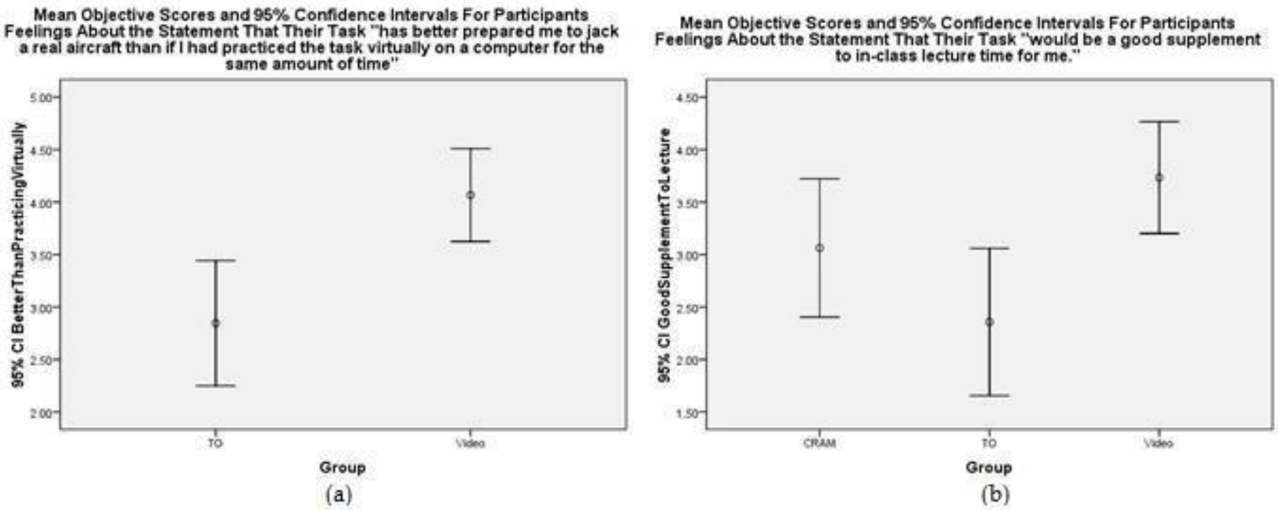

Fig. 4. Mean subjective scores and 95% confidence intervals for participants' feelings about the statement that their task: (a) "has better prepared me to jack a real aircraft than if I had practiced the task virtually" (b) "would be a good supplement to in-class lecture time for me"

ences in jacking experience based on block level. Post hoc analysis was done using Tukey's honestly significant difference (HSD) test. The test indicated that participants in course block 5 (M=4.57, SD=0.992) performed significantly better on the objective test than participants in block 3 (M=3.00, SD=1.069). There was no significant effect between either block with block 4 (M=3.87, SD=1.310).

Recognizing the influence of jacking experience and training group assignment, a two-way ANOVA was calculated on the objective scores of the participants to test for the interaction of jacking experience and training group assignment. The analysis was significant ($F(2,41)=3.413$, $p=0.043$) and Post hoc analysis using Tukey's HSD indicated the training group effect on the objective score is greater in the novice condition than the experienced condition. This means that in the novice condition, the video and CRAM group did even better against the TO than in the experienced condition.

Other variables from the demographic questionnaire were tested to determine if they had significant effects on the objective score. No effect on objective score was found based on a participant's age ($F(7,39)=0.489$, $p=0.837$), highest level of education ($F(4,42)=0.371$, $p=0.828$), rank ($F(3,43)=1.325$, $p=0.279$), choice in field assignment ($F(1,45)=0.745$, $p=0.393$), or the pairing of a participant with their preferred training method ($F(1,45)=0.503$, $p=0.482$).

In addition, the interaction between technological inclinations and training group assignments were tested to check for advantages or disadvantages that the technologically savvy or inept may have incurred while using the CRAM system. There was found to be no significant effect on objective score based on the interactions of training group and a participant's previous experience using interactive computer-based training systems (CBTs) ($F(2,41)=0.344$, $p=0.711$), comfort level with computers ($F(2,38)=0.324$, $p=0.725$), comfort level with technology ($F(6,35)=2.031$, $p=0.088$), affinity for learning new technologies ($F(3,38)=0.811$, $p=0.496$), frequency of video game play ($F(7,33)=2.004$, $p=0.084$), or frequency of computer usage ($F(3,39)=1.817$, $p=0.160$).

5.2 Subjective Scores

In addition to testing the objective improvement in hazard awareness for the three different training groups, the post-training opinions of participants were collected. A one-way ANOVA showed the effect of training method produced significant opinions in two scenarios. First, as shown in Fig. 4a, when asked to quantify their agreement (1-5, 1=strongly disagree, 5=strongly agree) with the statement that they believed their training method (video or TO) provided better preparation for jacking a real aircraft than virtual training, participants in the video training group ($M=4.07$, $SD=0.799$) agreed significantly more strongly ($F(1,26)=13.078$, $p=0.001$) than those in the TO ($M=2.84$, $SD=0.987$) training group. Note that this was a purely hypothetical question, asking participants to make judgments based only on their previous experience with virtual environments; these groups had no experience with CRAM. In addition, as illustrated in Fig. 4b, when asked to quantify their agreement (1-5, 1=strongly disagree, 5=strongly agree) with the statement that their training method (CRAM, video or TO) would be a good supplement to in-class lecture, participants in the video training group ($M=3.73$, $SD=0.961$) agreed significantly more strongly ($F(2,42)=5.228$, $p=0.009$) than those in the TO ($M=2.36$, $SD=1.216$) training group. No effect was seen for the CRAM group.

Training group assignment did not produce significant results in the subjective situations where participants were asked to: quantify their agreement with the statement that their training method could help them become more aware of hazards ($F(2,42)=0.652$ $p=0.526$) or was better than reading the TO ($F(1,29)=1.987$, $p=0.170$) or watching a video ($F(1,27)=0.057$, $p=0.813$).

5.3 User Satisfaction

The user satisfaction with CRAM was measured using the SUS. SUS scores range from 0 to 100 and are calculated by subtracting 1 each from the score of questions 1,3,5,7,9 and subtracting the score of questions 2,4,6,8,10 from 5. Then, the sum of the adjusted scores is multiplied by 2.5 to obtain the final SU value. The average SUS score for CRAM was 60.3; the implications of this score will be addressed in the discussion.

5.4 Discussion

The finding that training group had a significant effect on objective test score was expected. The hypothesis was that training using an interactive virtual simulation, such as CRAM, can improve hazard awareness over training methods such as reading the TO. That there is no statistically significant difference in groups that trained on CRAM versus watching a movie is a more interesting finding. It is clear from these findings that the participants can benefit from training tools that utilize visual technology over simple, written instructions, but not yet clear if those tools need to include 3D interactive simulations. Note that we did not have an opportunity to test the team-based components of the aircraft jacking procedure. It is possible that team communication and coordination as supported by CRAM could account for potential (hypothesized) future benefits of VR systems over pure video. It is also possible that adding better haptic feedback could improve these [28].

Not surprisingly, there is a significant difference between the scores of the novice and experienced groups. This distinction was controlled for by ensuring equal numbers of each experience group were in each training condition. One point to note is that if retention is this good for the chosen procedure, this task may have been the wrong one to choose as in most need of external practice methods. The simplest explanation for the difference in scores between blocks is that all students of block 5 had completed the jacking task, while students in blocks 3 and 4 had not. Less obvious is why the scores for students from block 5 are significantly greater than block 3 but not block 4. Although not showing a statistical difference does not mean that there is *no* difference, it may mean that block 4 subjects had more time to learn about hazards in general or had informally seen the procedure but not participated in it through formal classwork.

The subjective results are slightly more interesting in that only the video group expresses significantly stronger agreement over the TO when asked if their training method would be a good supplement to in-class lecture. One might suspect that the airmen are afraid or intimidated by the new technology, but looking into the demographic data collected from the participants, it is evident that this is not the case.

Table 1 Participants' comfort level with computers and technology. 1=Strongly Disagree, 2=Somewhat Disagree, 3=No Opinion, 4=Somewhat Agree, 5=Strongly Agree

	1	*2*	*3*	*4*	*5*
"I am comfortable using a computer"	2.1%	0%	2.1%	29.2%	66.7%
"I am uncomfortable using technology in general"	45.8%	29.2%	8.3%	0%	16.7%
"I enjoy learning new technologies"	0%	2.1%	10.4%	29.2%	58.3%

Table 1 shows that 66.7% of participants strongly agree with the statement "I am comfortable using a computer" and 29.2% more agree less strongly, but do not disagree with the statement. In total 95.9% agree at least to some extent. Similarly, 45.8% *strongly* disagree with the statement "I am uncomfortable using technology" and 29.2% more disagree less strongly, but do not agree with the statement. In total 75.0% disagree to some extent. Interestingly, even though only 45.8% *strongly* disagree that they're uncomfortable using technology, 58.3% *strongly* agree with the statement "I enjoy learning new technologies" and 29.2% agree to some extent, totaling 87.5% that enjoy learning new technologies.

Not only are the participants comfortable with technology but they use it to an incredible extent: 91.5% of them use a computer for social networking, 68.1% of them play computer games, 97.9% use email and 59.6% use some form of instant communication (Fig. 5). A simulation-based learning system, such as CRAM, can incorporate all of these features and teach hazards as well as a

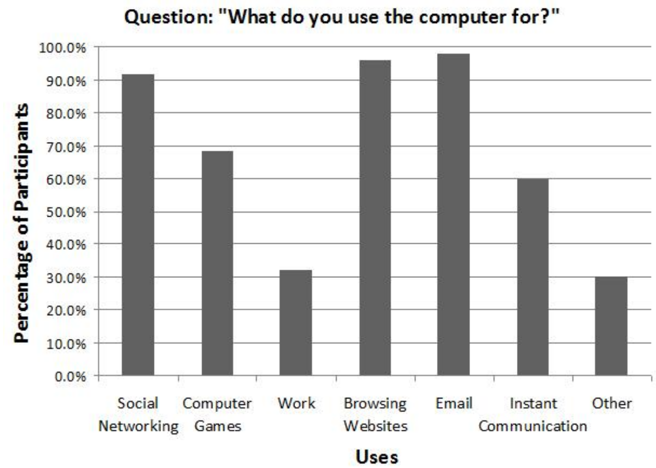

Fig. 5. Participants' computer usage

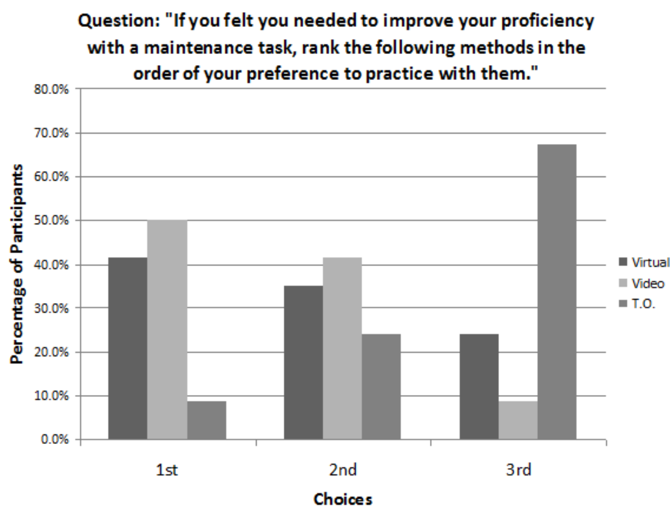

Fig. 6. Ordering of participants' preferred training methods

video does. Conceivably the participants did not realize that something that resembles the video games they play could teach them as well as something a video or reading from a manual. We have shown that it can.

Although some of the subjective scores point towards participants favoring the video, what should not be discarded is the unfavorable view of studying with the TO. As demonstrated in Fig. 6, when asked to put in order their preferred method of practice, *91.3%* chose either virtual training or watching a video, i.e. not the TO. Similarly, 76.1% chose one of the two technologies as their second choice. Two-thirds of the participants chose studying with the TO last. By implementing both training methods (which a simulation-based system can do - CRAM has embedded movie clips to go along with virtual steps), 91.3% of participants could train on their first choice of training method *and* would gain a better knowledge of hazards than if they had been asked to study using the standard written manual.

Finally, the average SUS score submitted by CRAM of 60.3 is promising but could be improved. Scores of 70 and over are generally considered to be the passing rate of usability for a system [29], and while this score indicates that CRAM was almost there, it could also provide insight into some of the negative feedback and lower than expected objective scores from the participants who used it. With a few simple improvements, a VR training system similar to CRAM could be a very usable system, which may lead to improved participant hazard learning. These needed improvements were collected from participants in their subjective questionnaires and will be discussed in the next section.

6 Conclusions and Future Work

We have described the design and evaluation of a prototype interactive software model to support technical training, as well as for use as an on-the-job task refresher aide. We demonstrated that both practicing virtually using CRAM and watching an instructor demonstrate a task in a video can improve a user's knowledge of hazards significantly over reading about them in a technical document such as the TO we provided. In addition, when given the choice of what they would like to practice on, participants overwhelmingly choose virtual practice or watching a video over reading the TO.

The suggested improvements to CRAM, made by the participants during the subjective questionnaire, are generally simple fixes such as the need for better navigation, strafe buttons (controls used to move laterally in first person shooter games), a more guided explanation about what is needed to complete each step – including more demonstration from the virtual coach – and better advertising of all of the components available (video, still images, etc.) for each step. Once these improvements are implemented, future versions of CRAM-like VR systems could use the SUS scores collected here as a baseline to confirm or disprove the hypothesis of improved usability and teaching capabilities.

Though many of the participants recognized in their subjective answers the utility of the CRAM system to allow them to practice safely, the most common complaint from those who did not value the CRAM function was that they would prefer their practice to be hands-on. This is to be expected from maintenance trainees but it may suggest that before having them fill out the SUS questionnaire it should have been more clearly stated that they should answer the questions with the understanding that when they respond to a statement such as "I think that I would like to use this system frequently" (question 1 on the SUS questionnaire) the choice they are making is between this system and another form of training that is not hands-on such as reading the TO or watching a video. Based on this choice between CRAM and the other two options, 41.3% of participants would prefer CRAM. For future evaluations using the SUS questionnaire, this distinction should be made clearer.

This evaluation captured more than just the objective and subjective utility of CRAM; it collected demographic data that gives insight into the future of training. Given the statistics presented in Fig. 5, it seems that rather than presenting CRAM as another CBT system, it could be better portrayed as a social networking site where trainees can play virtual simulations in which they must avoid hazards (or cause them, in a slightly modified version), can email instructors (through the wiki), and can have instant communication with other trainees during practice simulations, in both single- and multi-person tasks. As shown in Table 2, 74.4% of participants are playing video games 1-2 times per week (with 25.5% playing daily); that time could be at least partially allocated to playing the CRAM "game." Since 89.4% of participants are on the computer 1-2 times per week (with 68.1% on daily), we know that they can get to a computer twice a week, and thus could practice a task. A very high 93.6% of participants have internet access at home and therefore any web-based application is within their reach.

Table 2 Frequency of game play and computer usage: 1=Never, 2=Rarely, 3=1-2 times per month, 4=1-2 times per week, 5=Daily

	1	*2*	*3*	*4*	*5*
"How frequently do you play video games … that involve virtual environments?"	4.2%	12.5%	8.3%	47.9%	27.1%
"How frequently do you use a computer?"	0%	0%	10.4%	20.8%	68.8%

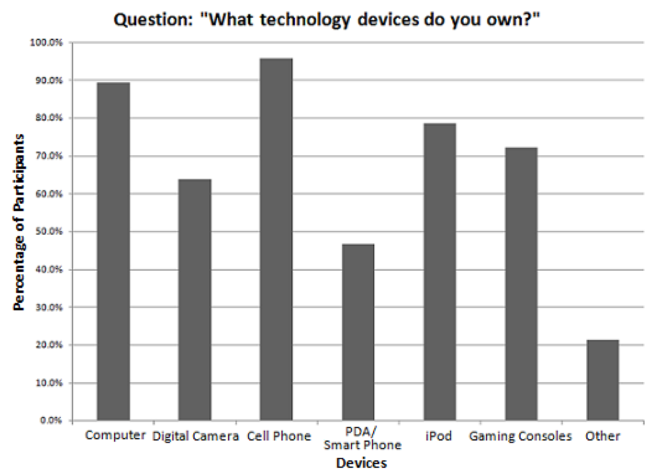

Fig. 7. Technology owned by participants

Apparently any sort of technology – even the simplest, such as a video – can be useful for and appreciated by our participant pool. The next step for VR systems, such as CRAM, would be to ensure even easier access by putting it on a platform that is appropriate for use on mobile devices. The capabilities on many PDA/Smart Phones could be utilized (Fig. 7 shows that 46.8% of the participants have them) or they could be made available as applications to be played on iPhones, iPods or Android devices (78.7% of participants have them). The main challenge in this system would be ensuring smooth 3D interaction in the much smaller 2D space of the phone. Finally, they could be extended to an even more game-like environment than the one we developed for CRAM.

The utility of a software-based, team training environment extends beyond what we have tested here. It could provide 24/7 instructional access (from the virtual coach), individual practice on group tasks (with virtual teammates) and training outside of the hangar/garage/workspace, all as an adjunct to hardware trainers and hands-on experience. We could not test the utility of all of these features here; we leave that to future work on VR training systems.

5 ACKNOWLEDGMENT

The authors wish to thank the personnel at the 82nd Training Wing (82nd TRW), Sheppard AFB, TX, including Ms. Gina Johnson, as well as instructor and student personnel assigned to the Air Force Fighter F-15 Aircraft Maintenance Apprentice Course (J3AQR2A333A025A). In addition, Grace Fong and Samantha Raja helped in the programming and creation of CRAM animations. Finally, 1st Lt. Dennis Riechman and Capt. Scott Pierce of the Air Force Research Laboratory and Patrick Vincent of Northrup Grumman Corp. for project management and logistics support in realizing the CRAM study. This work was supported in part by a grant from the Air Force Research Laboratory's Logistics Readiness Branch (AFRL/RHAL) under the Technology for Agile Combat Support (TACS) contract (FA8650-D-6546, Delivery Order #15).

REFERENCES

[1] S. Henderson and S. Feiner, "Evaluating the Benefits of Augmented Reality for Task Localization in Maintenance of an Armored Personnel Carrier Turret", *International Symposium on Mixed and Augmented Reality (ISMAR '09)*, pp.135-144, 2009.

[2] J.N. Templeman, L.E. Sibert, R.C. Page, P.S. Denbrook, "Pointman – A New Control for Simulating Tactical Infantry Movements," *IEEE Virtual Reality Conference (VR 200)7*, pp. 285-286, 2007.

[3] A. Raczynski and P. Gussmann, "Services and Training Through Augmented Reality," *Proc. European Conf. on Visual Media Production (CVMP '04)*, pp.263-271, 2004.

[4] W. Friedrich, "ARVIKA-Augmented Reality for Development, Production and Service," *Proc. Int. Symp. On Mixed and Augmented Reality (ISMAR '02)*, 2002, pp. 3-4.

[5] America's Army. Retrieved from http://www.americasarmy.com/

[6] Virtual Battlespace 2: US Army (VSB2). Retrieved from http://www.peostri.army.mil/PRODUCTS/VBS2/

[7] S.K. Ong, M.L. Yuan, and A.Y.C. Nee, "Augmented Reality Applications in Manufacturing: a Survey," *Int. Journal of Production Research*, vol. 46, pp.2707-2742, 2008

[8] J.W. Regian, W. Shebilske and J. Monk, "A Preliminary Empirical Evaluation of Virtual Reality as an Instructional Medium for Visual-Spatial Tasks," *Journal of Communication*, vol. 42, no. 4, pp. 136-149, 1992.

[9] J. Washbush and J. Gosen, "An Exploration of Game-Derived Learning in Total Enterprise Simulations," *Simulation & Gaming: An Interdisciplinary Journal*, vol. 32, no. 3, pp. 281-296, 2001.

[10] B.S. Bell, A. M. Kanar and S.W.J. Kozlowski, "Current Issues and Future Directions in Simulation-Based Training in North America," *International Journal of Human Resource Management*, vol. 19, no. 8, pp. 1416-1434, 2008.

[11] M.V. Veenman, F. J. Prins and J.J. Elshout, "Initial Inductive Learning in a Complex Computer Simulated Environment: The Role of Metacognitive Skills and Intellectual Ability," *Computers in Human Behavior*, vol. 18, no. 3, pp. 327-341, 2002.

[12] M.D. Williams, "A Comprehensive Review of Learner Control," *Proc. Ann. Conf. Assoc. Educational Communications and Technology (AECT '93)*, pp. 33, 1993.

[13] R. Thomas and E. Hooper, "Simulations: An Opportunity we are Missing," *Journal of Research on Computing in Education*, vol. 23, no. 4, pp.497-513, 1991.

[14] J. Swaak and T. de Jong, "Discovery Simulations and the Assessment of Intuitive Knowledge," *Journal of Computer Assisted Learning*, vol. 17, no. 3, pp. 284-294, 2001.

[15] D.J. Stipek, "Motivation and Instruction," *Handbook of Educational Psychology*, D.C. Berliner and R.C. Calfee, eds., New York: MacMillan, pp. 85–113, 1996.

[16] K.M. Sheldon and B.J. Biddle, "Standards, Accountability, and School Reform: Perils and Pitfalls." *Teachers College Record*, vol. 100, no. 1, pp. 164-180, 1998.

[17] H. Voke, "Student Engagement: Motivating Students to Learn," *Infobrief ASCD*, vol. 2, no. 28, 2002.

[18] S. Gerbaud, N. Mollet, F. Ganier, B. Arnaldi and J. Tisseau, "GVT: A Platform to Create Virtual Environments for Procedural Training," *Proc. IEEE Virtual Reality (VR '08)*, pp. 225-232, 2008.

[19] S. Babu, E. Suma, T. Barnes and L.F. Hodges, "Can Immersive Virtual Humans Teach Social Conversational Protocols?" *Proc. IEEE Virtual Reality (VR '07)*, pp.215-218, 2007.

[20] J.P. Chin, V.A. Diehl and K. Norman, "Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface," *Proc. ACM Computer Human Interaction (CHI '88)*, pp. 213-218, 1988.

[21] J. Nielsen, "Finding Usability Problems Through Heuristic Evaluation," *Proc. ACM Computer Human Interaction (CHI '92)*, pp. 373-380, 1992.

[22] J. Kirakowski and M. Corbett, "SUMI: the Software Usability Measurement Inventory," *British Journal of Educational Technology*, vol. 24, no. 3, pp. 210-212, Sept. 1993.

[23] J. Nielsen and T. K. Landauer, "A Mathematical Model of the Finding of Usability Problems," *Proc. Of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (INTERCHI '93)*, pp.206-213, 1993.

[24] The Federal Aviation Admin., "FAA Human Factors Workbench," http://www2.hf.faa.gov/workbenchtools/default.aspx?rPage=ToolList&subCatID=13. 2009.

[25] J. Brooke, "SUS: A Quick and Dirty Usability Scale," *Usability Evaluation in Industry*. P. Jordan, B. Thomas, B. Weerdmeester and McClelland, eds., London: Taylor and Francis, pp. 189-194, 1996.

[26] H. Kaufmann and A. Dunser, "Summary of Usability Evaluations of an Educational Augmented Reality Application," *Proc. HCI International Conference (HCII 2007)*, R. Shumaker, ed., LNCS 4563, Berlin: Springer-Verlag, pp. 660-669, Aug. 2007.

[27] S.T. Tullis and J.N. Stetson, "A Comparison of Questionnaires for Assessing Website Usability," *Proc. Usability Professionals Association Conference (UPA '04)*, June 2004.

[28] D. Vembar, A. Duchowski, S. Sadasivan and A. Gramopadhye, "A Haptic Virtual Borescope for Visual Engine Inspection Training," *IEEE Symposium on 3D User Interfaces, 2008 (3DUI 2008)*, pp.19-26, 2008

[29] A. Bangor, P.T. Kortum, and J.T. Miller, "An Empirical Evaluation of the System Usability Scale," *International Journal of Human-Computer Interaction*, vol. 24, no. 6, pp. 574-594, Aug. 2008.