Study
Note
2010-01

Usability of Wearable and Desktop Game-Based Simulations: A Heuristic Evaluation

John S. Barnett
U.S. Army Research Institute

Grant S. Taylor
University of Central Florida
Consortium Research Fellows Program

**United States Army Research Institute
for the Behavioral and Social Sciences**

May 2010

U.S. Army Research Institute
for the Behavioral and Social Sciences

Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:

Barbara A. Black   Michelle Sams

BARBARA A. BLACK, Ph.D
Research Program Manager
Training and Leader Development

MICHELLE SAMS, Ph.D.
Director

Technical review by

Martin Bink, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-ZXM, 2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926.

FINAL DISPOSITION: Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Study Note are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) May 2010	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) February 2009 – May 2009

4. TITLE AND SUBTITLE Usability of Wearable and Desktop Game-Based Simulations: A Heuristic Evaluation	5a. CONTRACT OR GRANT NUMBER
	5b. PROGRAM ELEMENT NUMBER 655803
6. AUTHOR(S) John S. Barnett (U.S. Army Research Institute) and Grant S. Taylor (University of Central Florida)	5c. PROJECT NUMBER D 730
	5d. TASK NUMBER 380
	5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: DAPE-ARI-IF 2511 Jefferson Davis Highway Arlington, Virginia 22202-3926	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 2511 Jefferson Davis Highway Arlington, Virginia 22202-3926	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Study Note 2010-01

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Subject Matter POC: Dr. John Barnett

14. ABSTRACT *(Maximum 200 words)*:

The use of simulators based on game software has the potential to deliver effective training. However, simulators with usability problems can interfere with training by presenting unwanted distractions. This report describes an assessment of the usability of a wearable computer system which has been designed to interface with a virtual environment and which can be used for simulator training. Usability of the wearable system was compared with that of a more common desktop interface in a game-based virtual environment. Eight evaluators conducted a heuristic usability evaluation of the wearable system and desktop interfaces. They identified 24 usability concerns with the wearable system and desktop interfaces, and the virtual environment. The majority of the concerns (46%) were with the virtual environment. Forty-two percent of the concerns were related to the wearable system, and the remaining 12% dealt with the desktop interface. However, when the frequency, impact, and persistence scores were aggregated into an overall score, the wearable system had the poorest usability. Eight of the ten greatest usability concerns were related to the wearable system. These data suggest that the virtual environment is more usable with the desktop interface than the wearable system.

15. SUBJECT TERMS

game-based simulations, usability, heuristic evaluation, immersive simulation

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	52	Ellen Kinzer Technical Publication Specialist 703-602-8049

Study Note 2010-01

Usability of Wearable and Desktop Game-Based Simulations: A Heuristic Evaluation

John S. Barnett
U.S. Army Research Institute

Grant S. Taylor
University of Central Florida
Consortium Research Fellows Program

ARI – Orlando Research Unit
Stephen L. Goldberg, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926

May 2010

Army Project Number	**Personnel and Training**
655803D730	**Analysis Activities**

USABILITY OF WEARABLE AND DESKTOP GAME-BASED SIMULATIONS: A HEURISTIC EVALUATION

EXECUTIVE SUMMARY

Research Requirement:

The use of simulators driven by commercial game software has the potential to deliver effective training. However, how closely the simulator must model the real environment has been a perennial question. The answer often depends on the type of tasks to be simulated. Tasks using motor skills may require very realistic models, whereas tasks employing cognitive skills can be less realistic. Training developers often assume that a simulator which is more realistic will provide better training, but this is not necessarily the case. A simulator which has physical similarity with the real system might not have the functional fidelity necessary to improve student's performance with the real system.

Many factors impact upon the training effectiveness of a simulator system, one of which is its usability. Simulators with good usability allow training time to focus on practicing skills to improve performance. Usability problems can interfere with training by shifting the focus from training objectives to the operation of the simulator technology.

Commercial wearable computer systems are available which have been designed to interface with a virtual environment and can be used for simulator-based training. While wearable interfaces have great potential for training, at present there is little information available on whether the wearable interface would be an efficient and cost-effective training delivery method.

One of the pieces of information which would help evaluate the cost-effectiveness of the wearable system is its usability, that is, how difficult is it to use compared to more common desktop interfaces. The research described in this report is the first step in evaluating the effectiveness of the wearable interface. It presents an evaluation of the usability of both the wearable interface and a desktop interface, as well as comments on the usability of the simulation environment.

Procedure:

Eight evaluators conducted a heuristic usability evaluation of two interfaces for the Game-based Distributed Interactive Simulation (GDIS). One interface was a common desktop keyboard and display, while the other was a wearable computer whose controls and displays mimicked equipment worn by a contemporary dismounted Soldier. The evaluators were given brief training on the operation of each interface and then directed through a scenario which prompted them to perform the majority of the capabilities of each interface. At the end of the scenarios, the evaluators were asked to identify potential usability concerns with the interface using a heuristic usability checklist as an outline for their evaluation.

Following the evaluations, the evaluators' concerns were aggregated and the evaluators were asked to rate the frequency, impact, and persistence of each usability concern. The ratings were analyzed and each concern was listed in order of its magnitude as a usability problem. In addition, the inter-rater reliability of the evaluators was measured and reported.

Findings:

Evaluators identified a total of 24 usability concerns with the wearable and desktop interfaces, and the GDIS software. The majority of the concerns (46%) were with the virtual environment software itself. Forty-two percent of the concerns were related to the wearable interface, and the remaining 12% dealt with the desktop interface. However, when the frequency, impact, and persistence scores were aggregated into an overall score, the wearable interface had the poorest usability. Eight of the ten greatest usability concerns were related to the wearable interface. These data suggest that the GDIS virtual environment is more usable with the desktop interface than the wearable interface.

Utilization and Dissemination of Findings:

These findings can be used to improve the usability of the virtual environment and the interfaces. The usability concerns (listed in the Appendix) can be addressed by designers and solutions may be found to improve the usability of the systems. For some concerns, the evaluators offer possible solutions to correct or improve usability concerns.

The findings can also be used as inputs for decisions on the costs and benefits of using different interfaces in conjunction with virtual environments as simulators for training. In its present state, the wearable interface is more difficult to use than the desktop interface. Several of the usability concerns suggest the wearable may produce negative training for some skills, such as aiming and employing the weapon and using the handgrip controls. Given the relative costs, in both time and money, of using the wearable interface versus the desktop interface compared to the amount of positive training produced by each interface, the results of this study suggest the desktop interface is a more effective training simulator than the wearable interface. However, correcting the usability concerns may change the relative training effectiveness of the interfaces. The costs associated with correcting usability problems can be included as part of the cost-benefit analysis to determine which systems would be the most effective and efficient for training.

The information in this report was provided to the sponsors at TPO Virtual on 14 January 2010.

USABILITY OF WEARABLE AND DESKTOP GAME-BASED SIMULATIONS: A HEURISTIC EVALUATION

CONTENTS

LIST OF TABLES

LIST OF FIGURES

USABILITY OF WEARABLE AND DESKTOP GAME-BASED SIMULATIONS: A HEURISTIC EVALUATION

At present there is considerable interest in the training community related to the use of computer games as simulators for training. Modifications of game engines can replicate realistic environments and avatars performing realistic tasks. The virtual environments and avatars can simulate the performance of certain tasks with enough realism that users can use them to learn, practice, and improve skills.

One of the issues with using computer simulations for training is that the simulation controls should be simple and natural enough that the trainee can focus on the training with minimal effort towards controlling the simulator itself. Also, simulators based on game engines can present tasks in ways that make sense while playing a game, but are significantly different from how the function is performed in reality. This gives rise to the question of whether the function is so far removed from reality that it gives no positive transfer of training, and perhaps instead may provide negative training transfer, whereby the trainee actually performs the real-world task worse as a result of their experience with the simulation.

Currently, the U.S. Army Research Institute (ARI) is working with two versions of a game based simulator for training Soldier tasks in a Military Operations in Urban Terrain (MOUT) environment. One version employs a desktop computer and uses the mouse and keyboard for controls and the computer screen for the display. The second version has the computer components, controls, and the display embedded into Soldier field equipment so that the simulator is wearable by the trainee. In the wearable version, the trainee's body movements are sensed by the computer and used as part of the input, controlling the Soldier's avatar in the simulated environment. The wearable simulator is expected to give the trainee more of a feeling of being immersed in the environment, which is thought to improve the training process.

As part of an evaluation of these two simulator versions for training, ARI conducted a heuristic usability evaluation of the two simulators to determine the ease of use of the two simulators' controls and displays. The purpose of the evaluation was to identify any function of the two simulators which might be especially difficult to use or might interfere with training.

This report describes the heuristic evaluation conducted with the two simulator versions. The results should prove beneficial to simulation engineers, as well as researchers and trainers utilizing similar systems.

Background

Heuristic Usability Evaluation

A heuristic evaluation is a means of considering a product or design to determine if it follows standard usability criteria. Its purpose is to find the most salient human-system interaction discrepancies (i.e., discrepancies between the design and accepted usability practices), either for the sake of the evaluation of a system prior to its implementation, or to guide the development team throughout the iterative design process. The process is designed to be easy to use, inexpensive, and be completed quickly, unlike more in-depth usability analyses which can be complex, expensive, and time consuming. Nielsen refers to it as "discount usability engineering" (Nielsen & Mack, 1994, p. 25).

In a heuristic evaluation, a number of evaluators interact with the product and evaluate it against a set of heuristic usability criteria (see Table 1). The usability heuristics serve as a framework for the evaluation. The heuristics were derived from a factor analysis of 249 usability problems (Nielsen & Mack, 1994). For a thorough description of the usability heuristics, see Nielsen (1993). There are a number of techniques for conducting a heuristic evaluation, including using specially designed checklists, or by having the evaluators interact with the product according to a typical user scenario.

Table 1. Definitions of common usability heuristics (Nielsen & Mack, 1994)

Visibility of system status The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.
Match between system and the real world The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.
User control and freedom Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.
Consistency and standards Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.
Error prevention Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.
Recognition rather than recall Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.
Flexibility and efficiency of use Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.
Aesthetic and minimalist design Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.
Help users recognize, diagnose, and recover from errors Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
Help and documentation Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Evaluators. Evaluators for heuristic analyses should have an understanding of usability concepts, as it would help them find more usability problems. Ideally, the best evaluators would have a knowledge of usability and be domain experts in the area being evaluated. However, such a combination is often rare. Nielsen and Mack (1994) suggest an understanding of usability principles is more important than domain knowledge in the selection of evaluators.

Nielsen and Mack (1994) found that a single evaluator will only find about 35% of the usability problems on average; therefore, they suggest using multiple evaluators. Their analysis of a number of heuristic usability projects suggests the relationship between the number of evaluators and the number of problems indentified is a curvilinear function that rises sharply from one to five evaluators, and tends to flatten out after that. Five evaluators should find about 75% of the usability problems, whereas ten should find about 85%. Their suggestion is to use at least three to five evaluators, with the number used depending on the costs of more evaluators for the particular project.

Method

Evaluators

Eight evaluators analyzed both the desktop and wearable versions of the simulator. Five were graduate students familiar with usability principles, two were applied psychologists, also familiar with usability principles, and one was a U.S. Army officer. Six of the evaluators had used the GDIS desktop system before, and three had used the wearable version before. The evaluators included seven men and one woman.

Apparatus

Software. The simulation software, GDIS (Game Distributed Interactive Simulation), is developed by the Research Network Institute (www.resrchnet.com) and is a modification of the Half-Life graphics engine developed by Valve. GDIS is a first-person shooter developed for the U.S. Army to be used for embedded training of Soldiers. The software allows for the use of customized interactive environments complete with buildings, terrain, vehicles, and friendly, enemy, and civilian characters (which can be fully automated or controlled by other live role-players). The current study has participants utilizing a virtual representation of the McKenna MOUT training site located in Fort Benning, GA.

Desktop Simulator. The desktop simulator is powered by a Dell XPS computer (2.67 GHz Intel Core 2 Duo processor, 3 GB RAM, NVIDIA GeForce 8800 GTX) with a 20 inch widescreen monitor. A standard keyboard and optical mouse are used for controls. Participants wear a set of headphones with an integrated microphone for radio communications. The controls used for the simulation are fairly typical of other PC-based first-person shooters (see Figure 1).

Wearable Simulator. The wearable simulator is powered by an ExpeditionDI system developed by Quantum3D (www.quantum3d.com). The system is comprised of a Thermite 1300 Tactical Visual Computer (1.4 GHz Intel Pentium processor, 1 GB RAM, ATI Mobility Radeon X300), which is worn on the back of a load-bearing vest. The computer presents the same GDIS simulation used on the desktop computers on a helmet-mounted eMagin Z800 SVGA OLED display. The system is fully self-contained and the user is not tethered to any external equipment, allowing them an unhindered range of motion. The user's movements are tracked via three tri-axis motion sensors connected to the head (helmet), simulated weapon, and thigh. The user controls their avatar through a combination of their own natural movements along with a joystick and series of buttons on the front handgrip of the simulated M4A1 rifle (see Figures 1 and 2). The user's head movements are used to control their view within the simulation, movement of the simulated weapon controls the position and aim of the weapon within the virtual environment, and the leg tracker is used to detect the user's posture (standing or kneeling) to adjust the avatar's position accordingly. The wearable interface is shown in Figure 2.

	Action	Desktop	Wearable
Movement	Move	W, S, A, D	Thumbstick
	View	Mouse Movement	Head Movement
	Aim	Mouse Movement	Weapon Movement
	Run	Shift	Thumbstick Threshold
	Kneel	Control (hold)	Kneel (leg tracker)
	Jump	Space	1
Actions	Weapon Select	Mouse Scroll	2
	Binoculars	Z (hold)	3 (hold)
	Flashlight	F	4
	Radio	\ (hold)	1+2 (hold)
	Compass Direction (map)	O	2+3
	Open Door	E	3+4
Firing	Fire	Mouse Click	Trigger
	Secondary fire (203, grenades)	Right Click	3+Trigger
	Reload	R	Remove/Replace Magazine
	Calibrate (wearable only)	---	1+2+3+4

Figure 1. Controls for the desktop and wearable simulators
Note: The numbered buttons on the wearable system indicate four buttons located on the front handgrip of the simulated M4A1 rifle (see Figure 2).

Procedure

Evaluators were welcomed by the experimenters and then given an overview of the evaluation. Since the evaluators were to judge the usability of the simulators based on heuristic principles, they were given definitions of the ten heuristic principles (Nielsen, 1993; Table 1) and allowed time to become familiar with them. Next, they were introduced to either the desktop or wearable simulator and briefed on the controls. The choice of which simulator the evaluators rated first was counterbalanced to eliminate any possible order effects affecting the heuristic ratings. They were allowed to practice with the controls until they became familiar with them. As part of the control familiarization, the experimenter asked them to perform a list of actions and prompted them on which control to use if necessary.

Once the evaluator was ready, they were asked to control their avatar in the simulated environment and follow a scenario where the avatar would perform certain actions. The experimenter guided them through the scenario by asking them to perform a set of tasks. When necessary, the experimenter would provide guidance on how to complete the action. The scenarios were designed to incorporate all of the functions the simulator could perform related to common military tasks such as movement, observation, target engagement, and communication.

Figure 2. U.S. Army Soldier wearing the ExpeditionDI wearable simulator interface (left). The front handgrip controls from the simulated M4A1 rifle (right).

When the evaluators completed the scenario, they were asked to record their evaluation on a questionnaire (see Appendix B). The questionnaire asked them to rate the simulator in reference to the ten heuristic principles on a 5-point Likert scale. The questionnaire also had space to comment on specific usability concerns of the simulator and make general comments. Reviewers were encouraged to report any and all usability concerns that they encountered, as this was the primary goal of this analysis.

When the questionnaire was complete, evaluators were taken to the second simulator and again allowed to familiarize themselves with the controls. The procedure used to evaluate the second simulator was similar to the first, except that a different scenario was followed. The two scenarios included the same tasks but in a different order, and the navigation around the environment was slightly different. The order of the two scenarios was also counterbalanced independently of simulator order. At the completion of the second scenario, the evaluator again completed a questionnaire identical to the first one.

Following the physical evaluation, the experimenters analyzed the feedback, aggregated similar comments, and identified a total of 24 unique usability concerns. Evaluators were then contacted by email and asked to complete an online survey as a follow-up to the evaluation. The survey presented the list of identified usability concerns and asked the evaluators to rate the severity of each concern. Evaluators were asked to rate the Frequency, Impact, and Persistence of each usability concern on a scale from 0 to 4, following guidance from Nielsen and Mack (1994). The Frequency ratings concerned how frequently the problem occurred, the Impact rated the difficulty a user would have to overcome the concern, and Persistence rated the extent to which the problem would extinguish over time (e.g. with more experience) or would persist even when the user gained experience with the system.

In addition to the severity ratings, evaluators were given the opportunity to comment on each usability concern and suggest methods to correct or improve the problem. They were also given space to list additional usability concerns not addressed that they felt were important. No additional usability concerns were submitted from any of the reviewers during this portion of the analysis.

Results and Discussion

Heuristic Ratings

The results of the reviewers' ratings of the usability heuristics for both systems are presented in Table 2. Ratings were made on a 5-point Likert scale on which higher values indicate better system performance. Due to the small sample size, a series of Wilcoxon Signed Ranks tests (the non-parametric equivalent of a t-test) were conducted to determine group differences between the two simulators within each of the usability heuristics. These tests show that the desktop system was rated significantly better on the Visibility ($Z = 2.232$, $p = .026$), Recognition ($Z = 1.980$, $p = .048$), and Error Recovery ($Z = 2.000$, $p = .046$) heuristics. The desktop system also received higher average ratings on all other usability heuristics, except for Match, though these results were non-significant.

Table 2. Mean reviewer ratings of the usability heuristics.

	Desktop	**Wearable**	**Wilcoxon *Z***	***p* (two-tailed)**
Visibility	4.42	3.25	*2.23	.026
Match	3.75	4.12	0.79	.429
Control	3.85	3.50	0.81	.414
Consistency	4.50	4.14	0.73	.461
Error Prevention	3.75	2.62	1.56	.119
Recognition	4.14	2.50	*1.98	.048
Flexibility	3.83	3.57	1.73	.083
Aesthetic	4.62	4.25	1.13	.257
Error Recovery	3.50	2.75	*2.00	.046
Help	4.00	2.66	0.44	.655

Note: Larger numbers indicate better system performance. * $p < .05$

Usability Concerns

A total of 24 unique usability concerns were identified from the collection of the initial heuristic analyses. Of these, 11 were pertinent to both systems, 3 were specific only to the desktop simulator, and 10 were specific only to the wearable system (see Appendix A for a complete list of all usability concerns with reviewer ratings, comments, and recommendations).

Inter-rater Reliability

The inter-rater reliability of the reviewers' responses for the usability heuristics as well as the usability concerns was computed using Krippendorff's alpha (Hayes & Krippendorff, 2007). This was done to determine the degree to which the reviewers agreed with one another. Krippendorff's alpha was used due to its unique ability to be applied to situations involving more than two raters in which ordinal data is obtained, which is the case with both of these situations. Inter-rater reliability was found to be fairly low for both the heuristic ratings ($\alpha = .1384$) and the ratings of the usability concerns ($\alpha = .3182$). Krippendorff's alpha computes inter-rater reliability on a scale of 0.0 to 1.0, and so these results indicate relatively little agreement among the reviewers on both measures.

Relative Usability of the Desktop and Wearable Simulators

Although the evaluation rated two simulator versions, the wearable and desktop simulators, the results highlight that there are actually three systems being examined: the wearable interface, the desktop interface, and the GDIS environment that both interfaces display. In fact, many of the concerns in the "both" category are actually concerns with the GDIS environment and the game engine used in the environment. Although, since it is impossible to employ the GDIS environment without an interface, or employ the interfaces without a virtual environment, it is probably most appropriate from a usability standpoint to consider the usability concerns of the desktop interface plus the GDIS environment together, and also to consider usability concerns of the wearable interface plus the GDIS environment jointly.

The evaluators found fewer usability concerns with the desktop interface than the wearable interface. The greatest number of concerns were categorized as "both" (pertaining to both the desktop and wearable systems), which suggests these were usability concerns with the virtual environment. As far as the number of usability concerns, 46% of the concerns were the result of the virtual environment ("both"); 42% were related to the wearable interface, and 12% concerned the desktop interface.

As well as the total number of usability concerns, it is important to consider the magnitude of the concerns. When ratings of Frequency, Impact, and Persistence are aggregated and averaged, the resultant score gives an indication of the relative magnitude of the usability concern. The aggregate scores for these usability ratings are shown in Figure 3.

In Figure 3, the light bars indicate usability concerns with the wearable interface, the medium gray bars are concerns with the virtual environment (rated as "both"), and the dark gray bars are concerns with the desktop interface. The height of the bars indicates the aggregate scores for Frequency, Impact, and Persistence. Thus, bar height indicates the relative severity of the usability concern, with higher bars indicating poorer usability.

A glance at Figure 3 shows the majority of the most severe usability concerns deal with the wearable interface. In fact, eight of the top nine concerns are unique to the wearable. The majority of the remaining usability concerns refer to both the wearable and desktop. Only three concerns relate to only the desktop interface, the poorest rated of these (i.e. poorest usability) ranks 15^{th} out of 24 concerns.

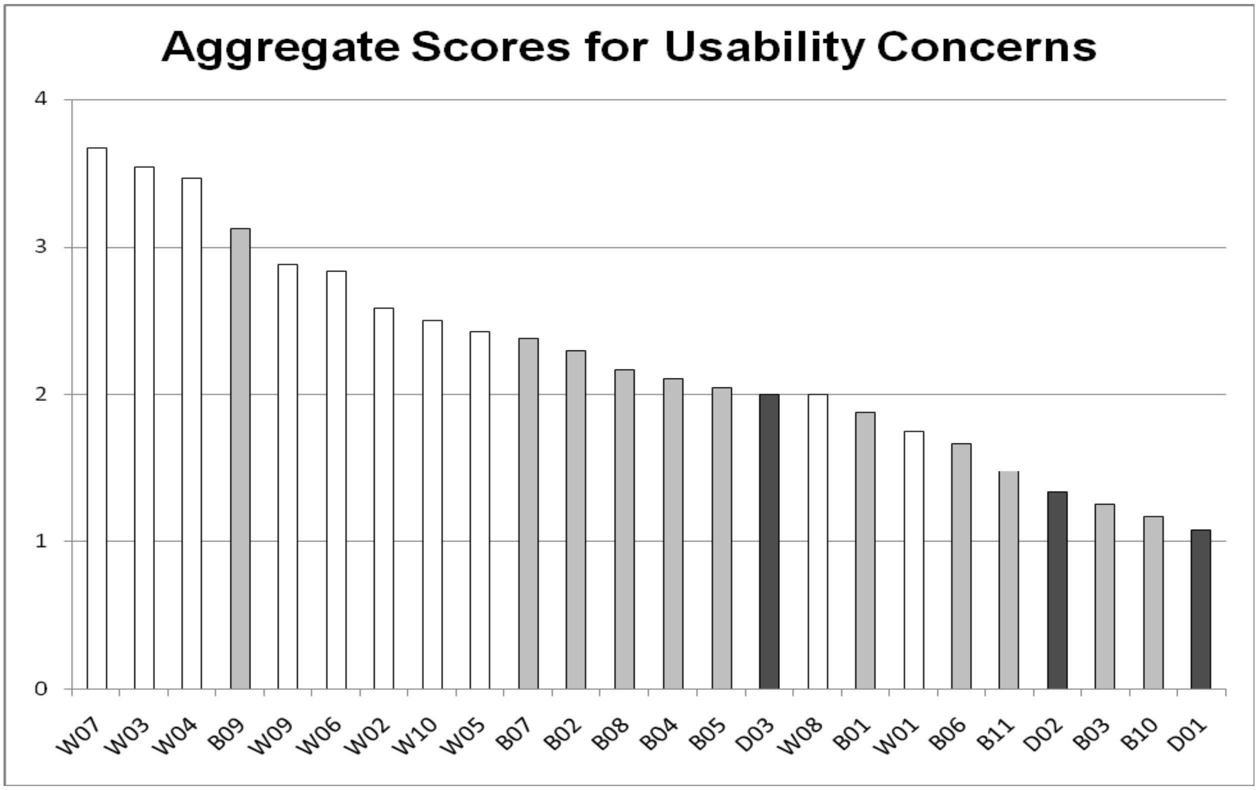

Figure 3. Aggregate scores for usability concerns. Light bars indicate concerns with wearable interface, medium gray bars show concerns with the virtual environment, and dark gray bars indicate concerns with desktop interface. The numbers shown are averaged ratings of frequency, impact, and persistence; higher numbers indicate more serious usability concerns.

Figure 4 further summarizes the results of these ratings by presenting the average rating of the Frequency, Impact, and Persistence associated with all of the usability concerns associated with each system. This shows that not only were there more usability concerns found related to the wearable system than the desktop, but that those related to the wearable system were consistently rated as being more detrimental than those related to the desktop system across every dimension.

From this it seems apparent that the desktop interface is easier to use than the wearable. When the usability concerns of the desktop interface are combined with the concerns of the GDIS environment, there are fewer and less severe concerns than the wearable combined with the GDIS environment. This suggests that Soldiers who use the GDIS environment for training would find it easier with the desktop interface than the wearable interface.

There are some factors which might shed light on the more favorable usability ratings of the desktop interface. For example, the desktop interface and controls are familiar to anyone who uses desktop computers (especially those who play PC-based games). Therefore, there may be considerable training transfer for anyone who is already familiar with the interface. Also, since the desktop interface is based on a commercial game engine, it follows many of the control and display conventions of commercial computer games, which have resulted from considerable user testing and natural evolution. In computer games, controls which are difficult to use typically do not enjoy commercial success and are abandoned. Thus, commercial games typically are more user-friendly.

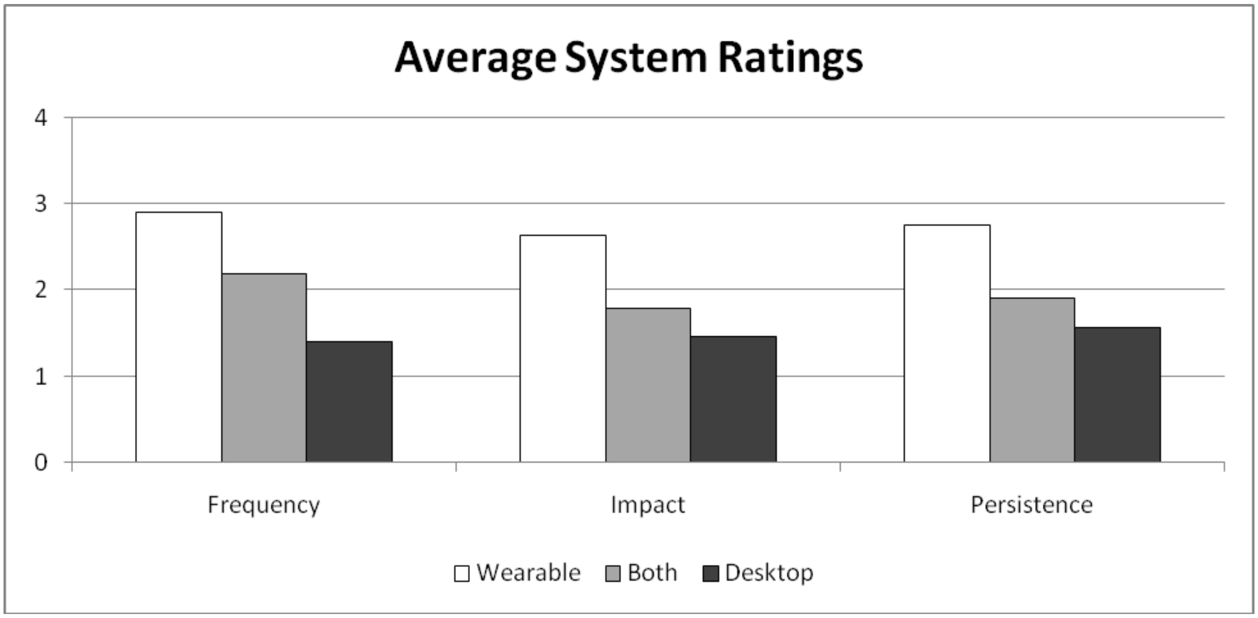

Figure 4. Average ratings of the frequency, impact, and persistence of the usability concerns related to each system. Larger numbers indicate more serious usability concerns.

Major usability concerns. Of the five most negatively rated usability concerns, four deal with the wearable interface, and one deals with the virtual environment. The most negatively rated concern involves the difficulty in using the simulated rifle control of the wearable interface. When the rifle control is raised into shooting position, the helmet, headphones, and helmet mounted display interfere with it and prevent the rifle from being positioned as a real rifle would be. When the rifle control is raised to firing position, the rifle graphic in the virtual environment is supposed to raise to firing position and allow the user to view the sights to shoot the target. However, the rifle graphic rarely is positioned so that the user can see the sights. To add to this concern, there is a cursor in the middle of the display that is used as a weapon sight in the desktop interface, but serves no purpose in the wearable system as bullets fire based on the orientation of the simulated weapon. Frequently, users will inadvertently use this cursor as the weapon sight, rather than the sights on the rifle graphic. This makes it very difficult for a user to hit a target with the wearable interface.

Our recommendation for correcting this problem is to provide an option in the GDIS software to disable the central crosshair when used with a wearable system; and then to increase the size of the area in which the system considers the simulated rifle to be "aimed" and in firing position, making it easier for the user to get a proper sight picture and shoot more accurately. If the difficulties aiming the weapon are a result of limitations inherent to using such a system (i.e. if the resolution of the accelerometer used to determine the position of the rifle control is too low), an additional button should be incorporated into the simulated weapon (possibly on the rear handgrip near the trigger) which puts the system into an "aim mode", which brings the virtual weapon up to a site picture regardless of where the simulated weapon is in the user's hands. Users would be instructed to press and hold this button as they are raising their simulated rifle to a firing position, allowing them to adjust their aim with the orientation of the simulated rifle, while still maintaining a proper site picture until the button is released.

Another concern with the wearable is the use of the four buttons on the forward hand grip (see Figure 2). Users find it difficult to remember the control sequences, and the button controls are unforgiving of slight delays in activating more than one button simultaneously. Frequently, users will press a two-button combination (i.e. 1+2 to activate the radio), but only one button will be recognized by the interface. This issue can be resolved, in part, by system administrators (researchers or trainers) limiting the number of functions available to the users, thus limiting the number of controls to memorize and limiting the number of controls requiring the use of multiple buttons in combination. However, as can be seen in Figure 1, the functions used in this exercise were relatively basic, and it would be difficult to design a training exercise that would not require the use of most or all of these functions. The primary problem is that the buttons are mapped arbitrarily to their functions (i.e. using button 3 for the flashlight is no more logical than using button 4), making it very difficult to remember which button controls a given function. This problem is avoided on the desktop system by mapping functions to keyboard keys that "make sense", such as using "F" for flashlight, making it much easier for users to learn and remember. This problem is somewhat unavoidable given the current control scheme with the wearable system, as there is no better way to map functions to the available controls. One improvement that could be made would be to avoid the use of the four handgrip buttons whenever possible through the implementation of more "natural" controls, such as an actual simulated flashlight mounted on the rifle, or a radio box that could be worn on the belt, with buttons on these objects to control their appropriate functions within the virtual environment. Of course, this additional hardware would increase the cost of the system, but should dramatically improve the user's experience.

Another concern rated in the top five deals with the difficulty of determining cardinal direction in the virtual environment with both simulators. Often, describing places and landmarks requires the use of cardinal directions, e.g., "the northeast corner of the building" or "the southeast room on the second floor." The evaluators noted that navigating within buildings seemed especially difficult. The only means of determining cardinal direction was by using a rudimentary compass in the overhead view display. Even with the compass, users seemed to have difficulty determining direction. This may have been because the compass display rotated as the user's view rotated, so that the compass, the user's view, and the desired direction became a complex mental rotation problem that many users find difficult. This concern may not be such an issue when the system is used by Soldiers, who are more accustomed to moving based on cardinal direction than our civilian reviewers, though our one reviewer with an Army background rated this concern similarly to the others.

One way to correct this issue would be to modify the way the map function works in the GDIS system. In the current design, pressing the "map" key once displays a small map at the top left corner of the screen, allowing the user to refer to it while still maintaining a first-person view of their surroundings. This small map has a compass rose on its bottom right corner, but it is too small to be usable by most users. Most users have to press the "map" key once again, which causes the map to fill the entire display, making the compass portion large enough to be usable. This makes it impossible to simultaneously view the compass direction and the environment first-hand, as most people would do when determining cardinal direction. This could be resolved by discarding the small map function and replacing it with a compass that can remain on the corner of the screen while the user still has a first-person view of the environment.

The fifth most negatively rated concern is the physical effort required to wear and operate the wearable system. Most of the evaluators showed signs of physical strain such as increased body heat and sweating after a few minutes in the wearable interface. Evaluators noted that users who are young and physically fit will have less of a problem with the physical effort involved. However, there is also the possibility of vertigo/dizziness (experienced by at least one evaluator) and simulator sickness, which affects some users regardless of their physical fitness. The physical effort required for the use of the system should not be of much concern for use in the training of Army Soldiers, who are accustomed to carrying similar loads regularly, however this must be seriously considered whenever a wearable system is implemented with users who may not be in such ideal physical condition.

Simulator sickness, on the other hand, is somewhat unavoidable, with some people being more susceptible to it than others, similar to motion sickness. The system does a good job of minimizing the likelihood of causing simulator sickness (there is no noticeable lag between movements and their display on the screen, and the head-mounted display does not occupy all of the user's vision), but users still need to be trained to recognize the symptoms early so that they know to take a break before their symptoms escalate. Other recommendations for reducing the possibility of simulator sickness can be found in Knerr, et al. (1998). A brief listing of these recommendations is shown in Table 3.

Table 3. Recommendations for Reducing Simulator Sickness (Knerr, et al., 1998)

- Participant's initial exposures to the virtual environment should be brief (10 – 15 minutes) with breaks of 5 – 10 minutes
- Avoid rapidly slewing field-of-view, collisions, and viewing objects at very short distances.
- Avoid depicting movement which is not under the participant's control.
- Initial sessions should be highly structured; avoid extensive "free play" where participants can do whatever they want.
- Periodically check the calibration of the system.
- Require participants to speak often. Changes in speech patterns may provide cues of developing sickness.
- The room should be cooler than normal, with fans for air circulation.
- If the participant is standing, provide a hard surface for hand contact.
- Have air sickness bags, or similar, available.
- Plan to retain participants until symptoms subside.

The highest rated concern with the desktop interface involves the proximity on the keyboard of the "shift" and "control" keys with the "windows" key. The "shift" and "control" keys are used in combination with the movement keys to run and crouch, and thus they are used regularly. This makes it easy to inadvertently press the "windows" key (usually located between "control" and "alt"), which causes the simulation to minimize and display the desktop with the Start menu open, effectively removing the user from the virtual environment. Restoring the environment window after an accidental press of the "windows" key frequently has negative consequences, such as unusual graphics colors or crashing the user's computer, requiring the restarting of either the GDIS software or the entire computer. Note that this is a problem inherent in the Windows operating system, and should not be considered a shortcoming of the GDIS software. This can be resolved relatively easily by system administrators by disabling the "windows" key functionality through the modification of the system registry files.

Although the five most negatively rated usability concerns were discussed here, all of the usability concerns found by the evaluators should be taken into consideration, with effort taken to resolve, or mitigate the negative effects of, these concerns. For the sake of brevity, the other usability concerns were not discussed here, but information on them can be found in the Appendix. See Appendix A for further discussion of all usability concerns found for all systems.

Training effectiveness. One issue to mention is the training effectiveness of wearable and desktop simulators. Theoretically, wearable interfaces provide more cues to the trainee in different modalities (spatial, kinesthetic, vestibular, etc.), which should improve training. However, the few research efforts which have addressed the relative training effectiveness of wearable and desktop interfaces found that head-tracked visual displays, body controlled movement, or a combination of the two can improve the performance of spatially-oriented tasks and acquisition of spatial knowledge, but beyond acquiring these few skills there are no clear differences in training effectiveness (Knerr, 2007).

Conclusion

It is important to stress that this research only evaluates the usability of the desktop, wearable, and GDIS environment; it does not directly address the relative training effectiveness or immersion of the systems. Even so, usability does have an indirect effect on training effectiveness. A training system which has poor usability will typically not train as well as one which is more usable. Usability concerns with training systems typically interfere with training because as more mental effort is spent dealing with training system problems, less mental resources are available for learning (Wickens, 1992). There is also the concern that users who must deal with usability shortcomings of training systems may be getting negative training, that is, the training system is eroding their real-world skills.

Future Research

The current research addresses the relative usability of the wearable interface versus the desktop interface within the GDIS virtual environment. Future research is needed to address the relative training effectiveness of these two systems. If one system is found to provide significantly better training than the other, then that system would be best for training. However, if neither system is superior, then the system which has better usability (i.e., the desktop interface) would be better for most training.

One factor to consider is that the two systems' effectiveness may vary with the tasks being trained. One system may be more effective for certain skills, while the other may be more effective with other sets of skills. Ideally, each system would be tested using the skills that are being taught in the field so that the measure of effectiveness would be valid.

To answer the question of which interface is better for delivering training, three pieces of information are required; usability, relative cost, and relative training effectiveness. With these data a cost-benefit analysis can determine the best training delivery at the lowest cost.

References

Knerr, B. W. (2007). *Immersive simulation training for the dismounted Soldier*. ARI study report 2007-01. Arlington, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Knerr, B. W., Lampton, D. R., Singer, M. J., Witmer, B. G., Goldberg, S. L., Parsons, K. J., & Parsons, J. (1998). *Virtual environments for dismounted Soldier training and performance: Results, recommendations, and issues*. Technical Report 1089. Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Nielsen, J. (1993). *Usability engineering*. San Diego: Morgan Kaufman.

Nielsen, J. & Mack, R. L. (1994). *Usability inspection methods*. New York: John Wiley & Sons.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*, 77-89.

Wickens, C. D. (1992) *Engineering psychology and human performance (2nd Ed.).* New York: HarperCollins Publishers.

Appendix A: List of Usability Concerns

This appendix includes a description of each of the usability concerns identified by the evaluators. The concerns are listed from most important (based on the mean of all evaluators' ratings for Frequency, Impact, and Persistence) to least important. Concerns that are specific to the wearable interface are preceded by the letter "W." Those that are specific to the desktop interface are preceded by the letter "D," and those that concern both interfaces are preceded by the letter "B." The numbers associated with each concern were used only to aid in organization prior to obtaining the reviewers' ratings, and are unrelated their relative importance.
Reviewers were presented with a list describing each usability concern, along with the following rating scales:

Frequency: How rare or common is this problem?
- 0- Very rare problem
- 1- Rare problem
- 2- Occasional problem
- 3- Common problem
- 4- Very common problem

Impact: How difficult is this problem for users to overcome?
- 0- I don't agree that it's a problem
- 1- Cosmetic problem - fix if time allows
- 2- Minor problem - low priority to fix
- 3- Major problem - high priority to fix
- 4- Usability catastrophe - imperative to fix

Persistence: Will this problem eventually go away or will it always be a problem?
- 0- Problem never should occur
- 1- One time problem - user learns after first mistake
- 2- Occasional problem - with training this problem will go away
- 3- Frequent problem - will often frustrate user
- 4- Very frequent problem - will constantly frustrate user

The reported results include the mean, mode, and standard deviation (SD) for each rating of each usability concern. The mean is reported as a measure of the average of all reviewer ratings. The mode, by describing the most common response, identifies which response most reviewers agreed upon (no rating was bimodal). The standard deviation is a measure of the variance within the responses, and is reported as a means of identifying those ratings reviewers rated more consistently, and those that received more mixed ratings (a higher standard deviation indicates less consistency among the reviewers).

Concern W07

Description. Aiming is difficult in the wearable system due to problems calibrating the weapon, the unnecessary crosshair at the center of your vision, and interference between the weapon and the HMD preventing holding the weapon in a proper firing position.

Table A-1. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern W07

	Frequency	Impact	Persistence
Mean	3.75	3.5	3.75
Mode	4	4	4
SD	0.46291	0.755929	0.46291

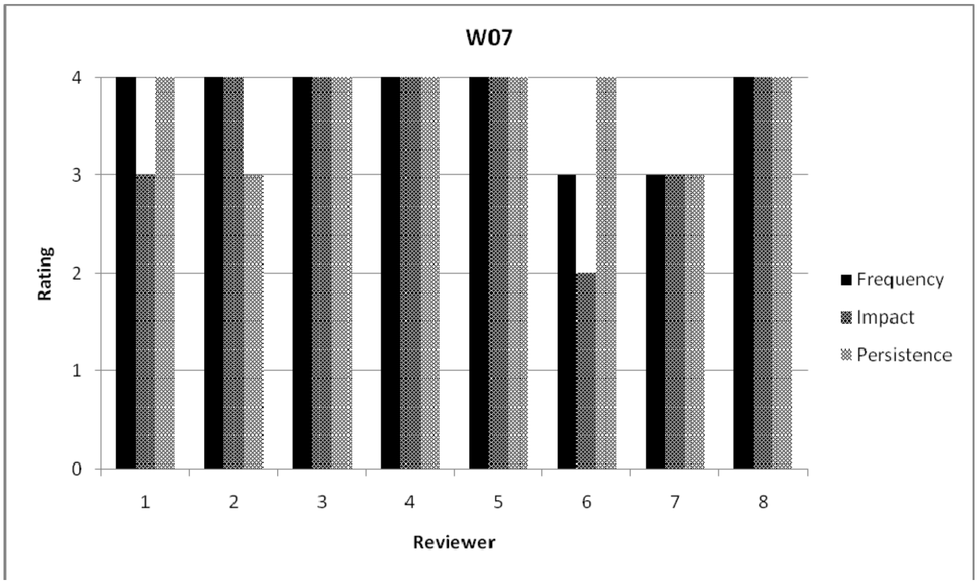

Figure A-1. Evaluators' Ratings for Concern W07

Evaluators Comments

- General comments
  - Plus weapon weight.
  - It is very difficult to hit a target in the current configuration. This makes it demotivating for users and also there is the potential for negative transfer of training.
  - big issue with wearable...too confusing a sight picture
  - Agreed. The system perhaps can be fine tuned and calibrated better.
- Recommendation
  - Remove central crosshair from the wearable system. Improve the hardware to improve the calibration process and prevent the weapon from "drifting", allowing the virtual weapon to come up to where it should be so the user can look down the sights. If hardware limitations do not allow this, designate a button on the weapon to "Aim", so that when it is pressed the virtual weapon is raised and the user can look down the sights, and still move the weapon to adjust their aim while maintaining a sight picture.
  - Redesign
  - This is a hardware possibly software issue.

Concern W03

Description. Actions in the wearable system that require the use of the 4 handgrip buttons (especially those that require combinations of buttons) are difficult to remember and require additional training.

Table A-2. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern W07

	Frequency	Impact	Persistence
Mean	3.75	3.5	3.375
Mode	4	3	4
SD	0.46291	0.534522	0.744024

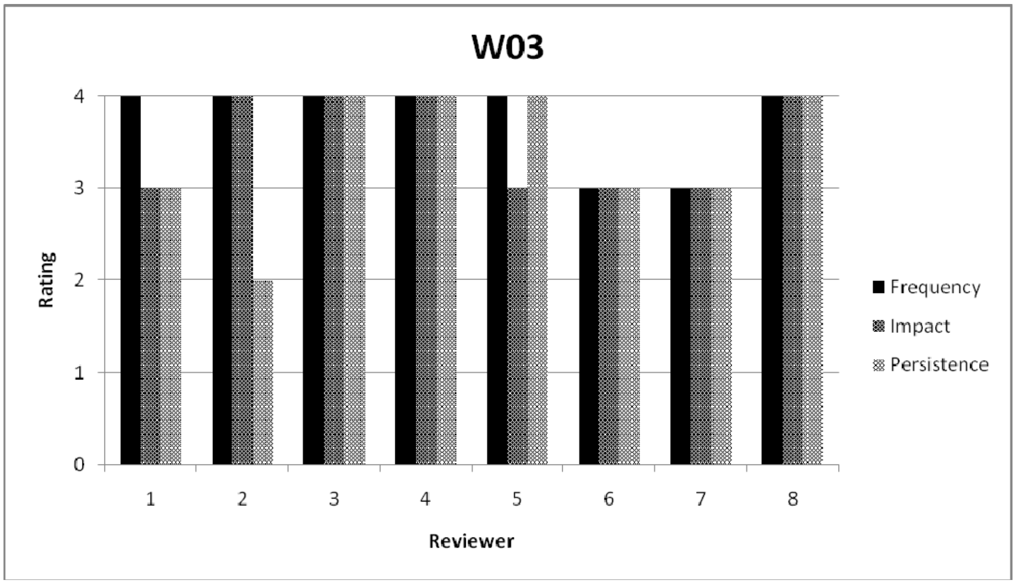

Figure A-2. Evaluators' Ratings for Concern W03

Evaluators Comments

- General comments
  - Controls are hard to remember.
  - Yes, big problem. The finger splay required to operate the buttons, for smaller hands, makes it nearly impossible to have dexterity or strength.
  - more practice time would probably cure this
  - Difficult to remember during moments requiring quick reactions
  - Agreed. The unforgiving nature of the simultaneous presses also is frustrating.
- Recommendation
  - Need new help display which is limited to available controls and easy to read.
  - Create more "natural" controls (i.e. a simulated radio worn on the hip, pressing a button on the simulated radio to activate the radio function rather than an arbitrary handgrip button will more accurately model reality and be easier to remember).
  - Control layouts for the handgrip should consider the relative strength and dexterity of the fingers, e.g. the ring and little fingers are not as strong or dextrous as the others, therefore third and forth buttons should not be used for common controls (like recalibrate).
  - Major control revamp recommended.

Concern W04

Description. The handgrip controls on the wearable system are difficult to use when pressing buttons in combination, requiring exact timing for combination presses.

Table A-3. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern W04

	Frequency	Impact	Persistence
Mean	3.5	3.25	3.625
Mode	4	4	4
SD	0.755929	0.886405	0.517549

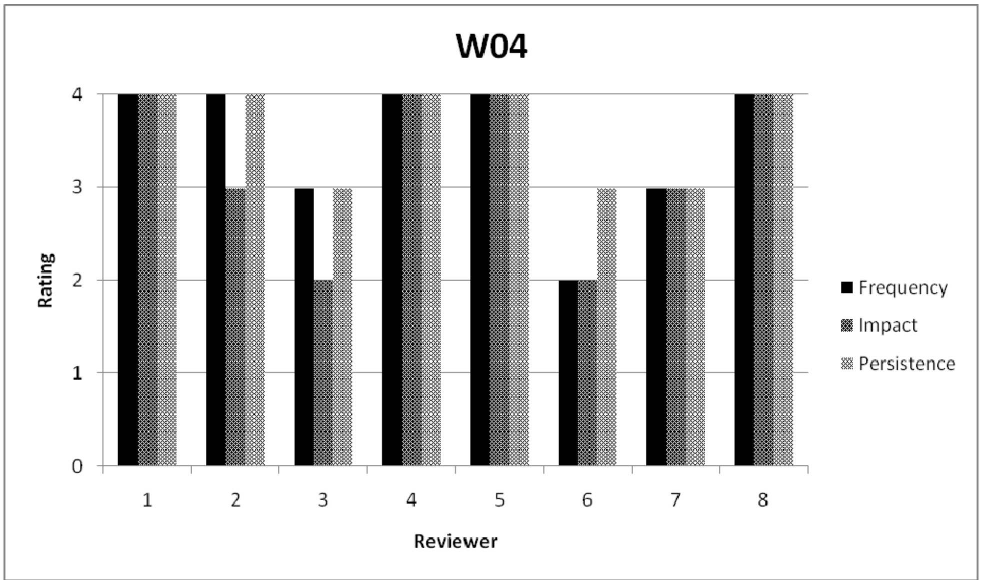

Figure A-3. Evaluators' Ratings for Concern W04

Evaluators Comments

- General comments
  - This is a major problem.
  - Major problem.
  - Concur
  - This has already been addressed in another question.
- Recommendation
  - Create enough "natural" controls (as described above) to avoid the need for combination button presses. If button combinations are unavoidable, allow for more error when determining whether buttons were pressed simultaneously, or have a separate "Shift" button that can be held down (rather than pressed simultaneously) to alter the function of the 4 handgrip buttons.
  - Try to loosen timing restrictions.
  - ask Soldiers what works best
  - CHANGE IS NEEDED. A shift key system or relaxing of the simultaneous nature of the button presses would be welcome

Concern B09

Description. It is difficult to determine cardinal direction.

Table A-4. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern B09

	Frequency	Impact	Persistence
Mean	3.25	3.125	3
Mode	4	3	2
SD	0.886405	0.834523	0.92582

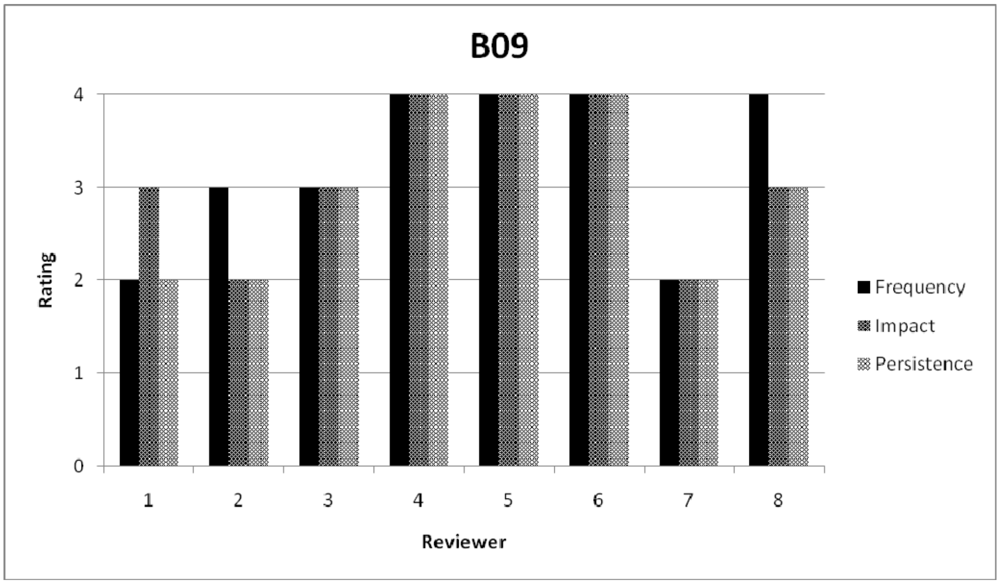

Figure A-4. Evaluators' Ratings for Concern B09

Evaluators Comments

- General comments
    - I agree that it is much more difficult to determine cardinal direction in the game than is in the real world. The only solution is to use the map/compass. This problem will be greatly reduced with practice.
    - This was a problem mainly because of the color and legibility of the compass on the mini-map. By maximizing the map, one could read it clearly but having a user bring up the full map every time he or she wants to check a direction is frustrating.
    - It's also easy to get lost in buildings.
    - Pop-up compass is basically worthless and a lot of time spent trying to use it.
    - In the wearable version it's hard to see the map.
    - This I agree with, especially on the wearable system. The compass is super bright and gets lost.
- Recommendation
    - Provide a compass in a corner of the display (on both the desktop and wearable system) that is separate from the map and can be toggled on and off.
    - No change
    - Change the color of the compass to stand out more if possible, or give the compass portion of the map a small solid color background to be laid up against.
    - More detail on buildings (posters, signs, different paint schemes, furniture, etc.) may allow users to more easily differentiate buildings/rooms. Also, a better heading system (compass, etc.) is necessary, and it should be always displayed versus user selected.
    - Recommend continuous "heads-up" compass that shows cardinal directions whichever way you are facing
    - Adjust the contrast or put a 3D arrow or other guide marker on the top of the screen.

Concern W09

Description. Physical discomfort in the wearable system. Physical strain, heat/sweating, simulator sickness, claustrophobia, headache are all common and/or possible.

Table A-5. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern W09

	Frequency	Impact	Persistence
Mean	3	2.875	2.75
Mode	3	2	3
SD	1.309307	0.834523	1.035098

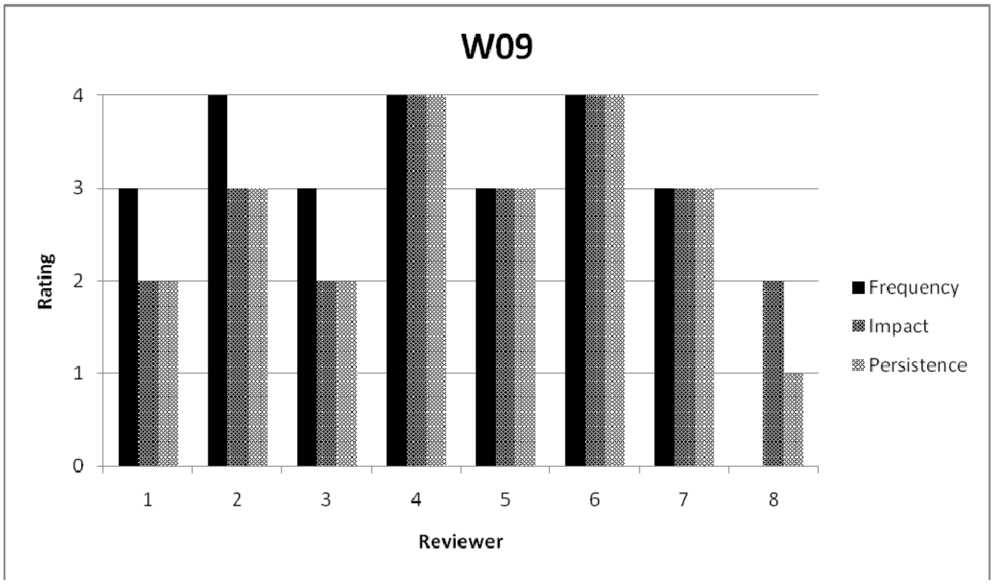

Figure A-5. Evaluators' Ratings for Concern W09

Evaluators Comments

- General comments
  - The physical strain is far less of a problem for young, fit Soldiers than for us civilians.
  - It was hot after a while. Other than that I think most of my discomfort was due to the fact that I am grossly out of shape.
  - This requires that users are watched to ensure they do not become over-strained. Users may not notice when they are over fatigued if they are immersed in the simulation.
  - I did not experience this personally though people who do have a problem with this certainly limit the effectiveness of the system.
- Recommendation
  - Lighten all aspects of the system to the extent possible. Use a mesh rather than solid fabric for the vest to avoid trapping body heat. Train users to recognize simulator sickness symptoms, and how to quickly exit the simulation if they experience any of them (flipping up the HMD seems to be the best/easiest way to do this quickly).
  - Look on commercial and mil industry markets for something more useable
  - Medication? smaller movements? Brightness/contrast? Not sure what the trigger is.

Concern W06

Description. The thigh tracker requires too specific of an angle to cause the avatar to kneel.

Table A-6. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern W06

	Frequency	Impact	Persistence
Mean	2.75	2.75	3
Mode	4	3	3
SD	1.28174	1.164965	1.069045

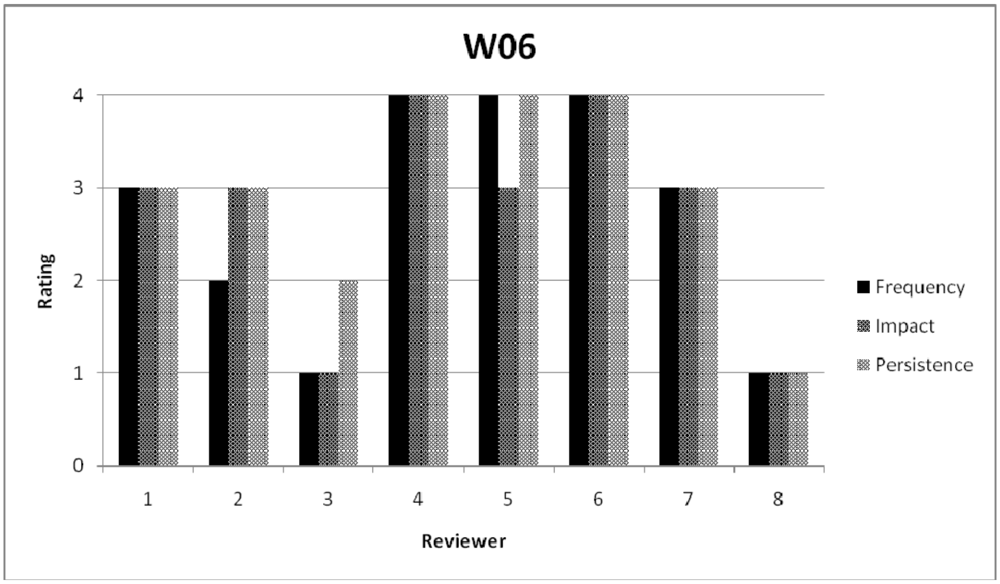

Figure A-6. Evaluators' Ratings for Concern W06

Evaluators Comments

- General comments
  - Did not experience this short of one time.
  - hard to kneel down
  - People with knee problems will find that very hard to use.
  - I did not experience this, perhaps a calibration issue?
- Recommendation
  - Increase the range of angles the thigh tracker accepts as being "kneeled", and have this parameter modifiable by the user/system operator.
  - Redesign
  - See how many really had this problem.

Concern W02

Description. A controls reference (similar to the sheet provided for the desktop) is needed for the wearable system.

Table A-7. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern W02

	Frequency	Impact	Persistence
Mean	3	2.625	2.125
Mode	4	2	2
SD	0.92582	1.06066	0.991031

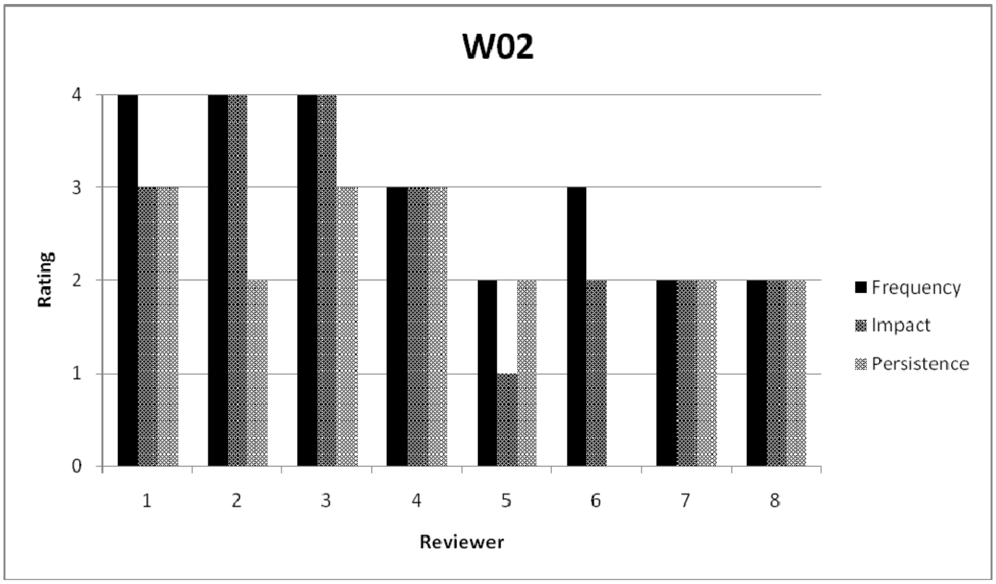

Figure A-7. Evaluators' Ratings for Concern W02

Evaluators Comments

- General comments
  - Controls are hard to remember.
  - The problem I see here is how to implement one. This may be something useful to use at the beginning of training in conjunction with additional training.
  - Yes, it's frustrating.
  - How would you use this once the HMD is on?
  - I would agree a control sheet would be valuable
  - Another good idea.
- Recommendation
  - The controls reference system proposed above should be implemented for the wearable system as well.
  - Need new help display which is limited to available controls and easy to read.
  - Pop up window, or perhaps a reference on the floor.

Concern W10

Description. The wearable system is prone to brief freezes, low frame rate, and lag.

Table A-8. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern W10

	Frequency	Impact	Persistence
Mean	2.5	2.375	2.625
Mode	3	2	3
SD	0.92582	0.916125	0.916125

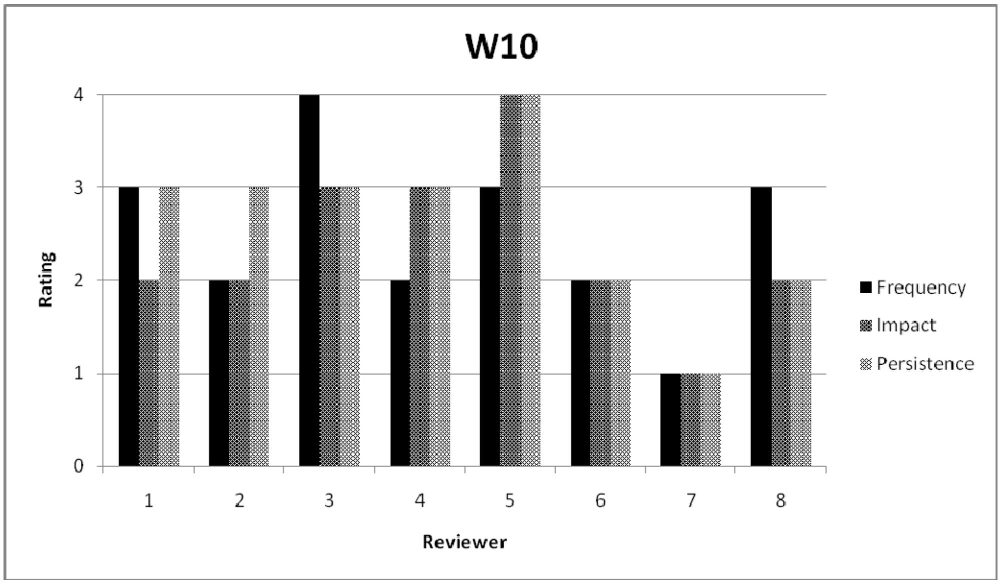

Figure A-8. Evaluators' Ratings for Concern W10

Evaluators Comments

- General comments
  - Unfortunately this is inherent in the system.
  - It was frustrating to me as a gamer, maybe due to my expectations about how a game should run. If my games at home ran like that, I would immediately fix it lest I drive myself insane.
  - Didn't notice when I was in it, but this would definitely be a problem if it occurs at all.
  - Lag and slowdowns interfere with both cognitive and physical actions and cause considerable frustrations for users.
  - This is true, but with better hardware this will likely be managed.
- Recommendation
  - System hardware should be improved, or the graphics quality should be lowered to improve system speed.
  - Better hardware

Concern W05

Description. Combining natural and artificial controls in the wearable simulator creates confusion. The fact that some of the controls are natural movements makes those that are unnatural (i.e. moving with the thumbstick) seem awkward and unnatural by comparison.

Table A-9. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern W05

	Frequency	Impact	Persistence
Mean	2.625	2.125	2.5
Mode	2	2	2
SD	0.744024	0.834523	1.069045

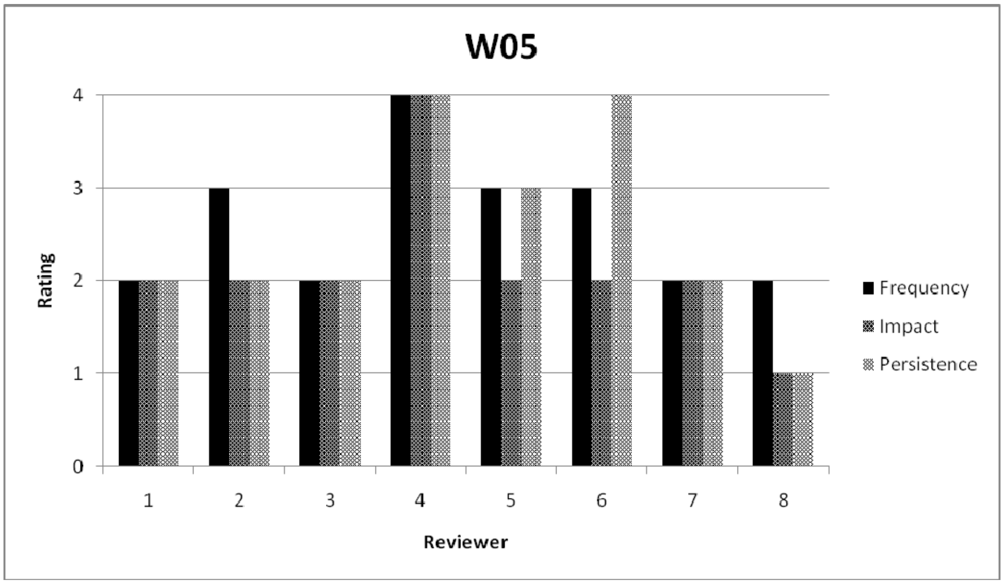

Figure A-9. Evaluators' Ratings for Concern W05

Evaluators Comments

- General comments
    - This is something that is inherent in the design of the simulator. I do not see a fix.
    - Yes, major problem.
    - yes - this was the greatest problem I had
    - True. But this is what we have now. It's not too far a stretch, and people who play videogames will find the control scheme familiar.
- Recommendation
    - Create as many "natural" controls as possible, although some controls (such as throwing grenades) seem incapable of being controlled via natural movements.

Concern B07

Description. The weapon selection process is difficult, especially the speed required to select a highlighted weapon in order to activate it.

Table A-10. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern B07

	Frequency	Impact	Persistence
Mean	2.625	2.25	2.25
Mode	2	3	3
SD	1.06066	0.886405	0.886405

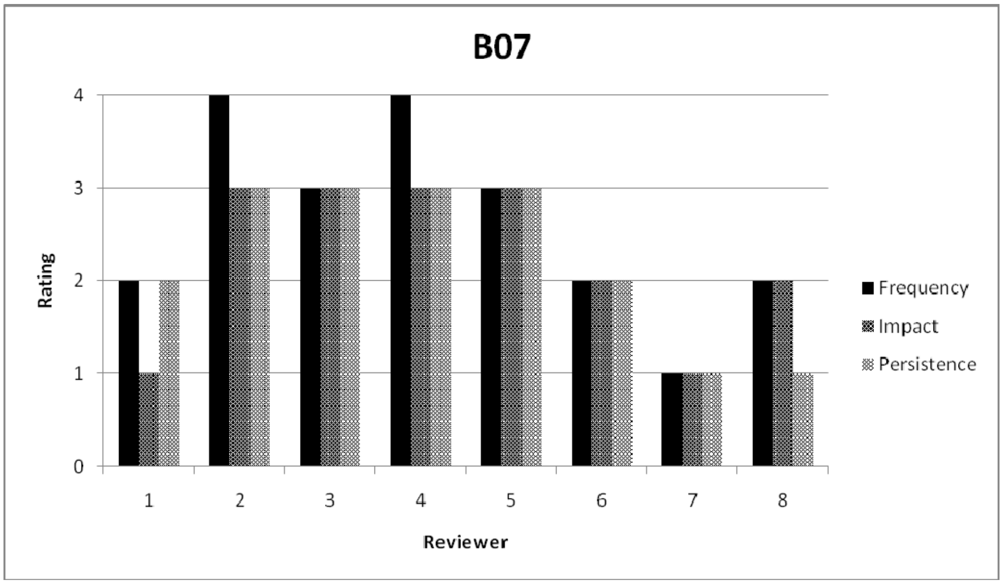

Figure A-10. Evaluators' Ratings for Concern B07

Evaluators Comments

- General comments
    - This problem can be reduced if users are trained to use the number keys to select the general weapon category.
    - Wearable seemed worse than the desktop. Wearable felt like I only had a split second to change, maybe due to the extra motor skills needed to switch as opposed to using the mouse.
    - only a problem in wearable set
    - This can also be altered. Or simply have hotkeys for each weapon i.e. 1 for gun, 5 for grenade, etc.
- Recommendation
    - Keep the current system of scrolling through the weapon list with the mouse scroll wheel (or handgrip button on the wearable system), but allow more time to select the currently highlighted weapon before exiting the weapon menu. Can also require pressing the scroll wheel to select the highlighted weapon (desktop only), so that left and right click can still be used to fire the currently held weapon while the weapon menu is open.
    - Change the training.
    - Fewer weapons choices and using keys to select weapons versus the mouse wheel may alleviate this problem somewhat.
    - You can select weapons via the number keys on the desktop version. Probably a good idea to let the users know.
    - Another easy fix.

Concern B02

Description. Additional instruction is needed for some users.

Table A-11. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern B02

	Frequency	Impact	Persistence
Mean	2.5	2.25	2.125
Mode	2	2	2
SD	0.534522	0.707107	0.64087

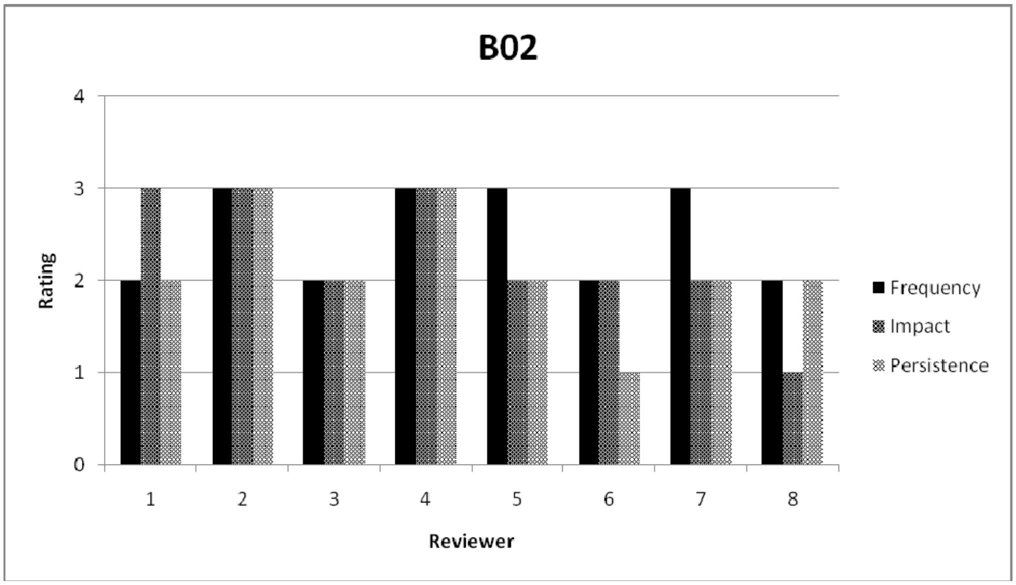

Figure A-11. Evaluators' Ratings for Concern B02

Evaluators Comments

- General comments
  - Some users, especially those unfamiliar with the game engine may require reminders on the operation of the system.
- Recommendation
  - Create a self-paced familiarization phase, so that users requiring more time to become comfortable with the controls can have it without frustrating those that don't.
  - I was in that group. The wearable set was difficult to master
  - This is a training tool, so people are expected to be trained on the tool as well, I don't think this is a real problem.

Concern B08

Description. Throwing grenades accurately is difficult.

Table A-12. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern B08

	Frequency	Impact	Persistence
Mean	2.5	1.875	2.125
Mode	3	3	1
SD	0.755929	1.125992	1.125992

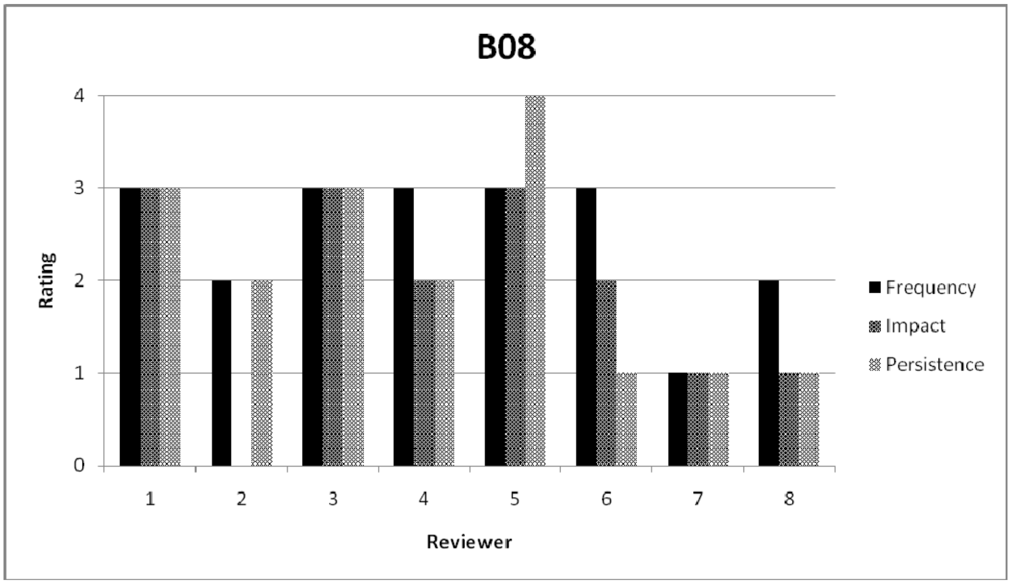

Figure A-12. Evaluators' Ratings for Concern B08

Evaluators Comments

- General comments
  - I do not think that throwing grenades is unrealistically difficult.
  - This may be because the grenades do not fly out from where the user expects them to.
  - concur, even if you roll them
  - At first but with practice it gets easier
  - Another aspect of training. You get better at it.
- Recommendation
  - When grenade/smoke is held, show the point on the ground where it will land when thrown from the current position. Although this is unrealistic information to provide, it compensates for the lack of tactile feedback necessary to accurately throw an object a specific distance.
  - Do not make any changes.
  - safety statement so you don't frag yourself or a buddy
  - An "auto-aim" feature can be implemented for grenades if need be.

Concern B04

Description. Some of the display information is confusing/unnecessary/unrealistic (team affiliation, unnecessary Suit value, crosshair provides an unrealistic advantage by showing exactly where shots will land, ammo numbers are unlabeled).

Table A-13. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern B04

	Frequency	Impact	Persistence
Mean	2.75	1.428571	2.142857
Mode	3	1	2
SD	1.035098	0.534522	0.690066

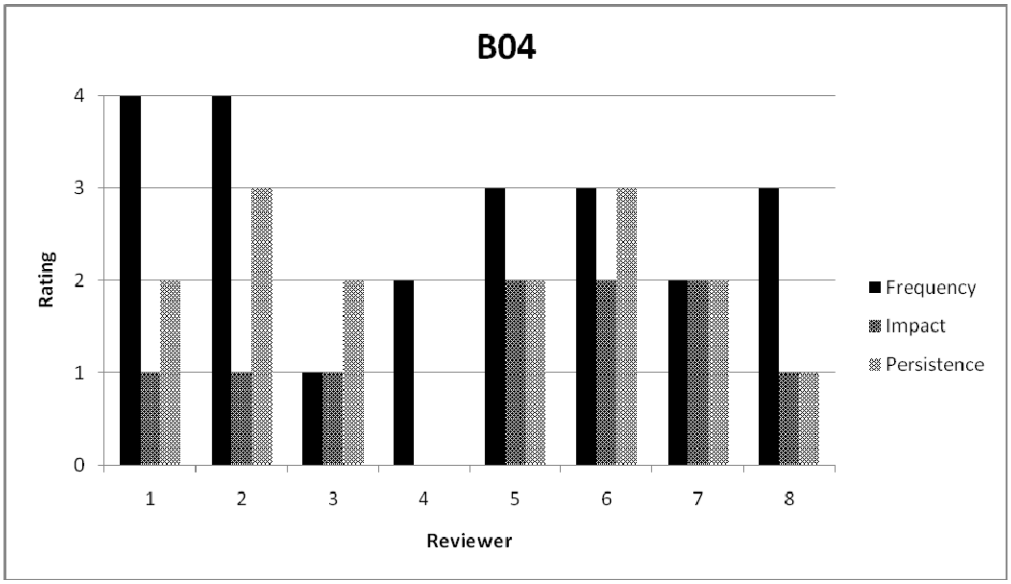

Figure A-13. Evaluators' Ratings for Concern B04

Evaluators Comments

- General comments
  - There are multiple issues here: irrelevant info (from a training standpoint), crosshairs, and lack of labels on the ammunition counts.
  - I think these issues would be specific to whatever was being trained. For example, when in this usability study, I didn't notice "team affiliation" or a "suit valve" nor did I realize that I should be tracking "ammo numbers". But if these were elements needed for training they may be important. So I'm not really sure how valid my rating is for this question--depends on the intended use.
  - Even though the displays are peripheral, they still tend to occlude part of the environment.
  - Yes, I think misinformation, especially in overload moments, is very distracting.
  - It can be a little confusing at first using the crosshairs to throw grenades and not using them to aim the gun.
  - This is also not a real concern, this is a training tool. The crosshair "advantage" is removed once the wearable system is used.
- Recommendation
  - Provide an option to remove each aspect of system information individually (Health, Suit, Endurance, crosshair, team affiliation, rounds in current magazine, reserve rounds, and secondary fire rounds), so that they be disabled when the system administrator believes they should be. Also label ammunition numbers.
  - Remove some of the irrelevant information. Add labels to the ammo counts if it is easy to do.
  - Remove unneeded displays.
  - This one is okay.

Concern B05

Description. It is easy to accidentally fire the weapon on both systems. There is a safety available on the weapon in the wearable, but not on the desktop.

Table A-14. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern B05

	Frequency	Impact	Persistence
Mean	2.25	1.75	2.125
Mode	1	3	3
SD	1.488048	1.164965	0.991031

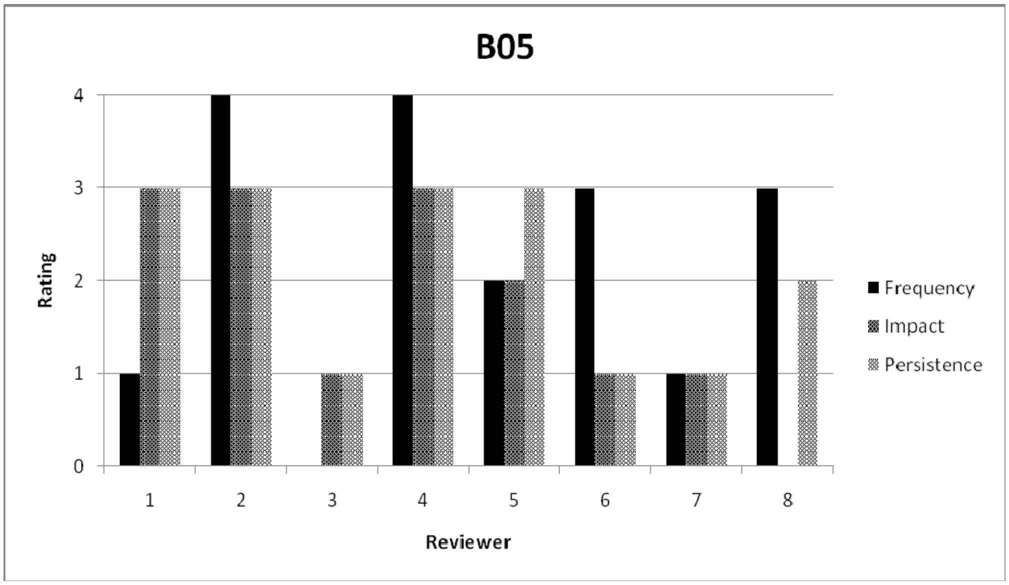

Figure A-14. Evaluators' Ratings for Concern B05

Evaluators Comments

- General comments
    - I'm not sure how valid a rating of "impact" or "persistence" is for the usability study. It was not a big problem in the confines of checking out the usability, but would obviously be a problem if you didn't want to fire on someone in an actual scenario. Then depending on the density of the forces, it could be a real problem. And if there is a need for quick response and for accuracy of firing, it would be a major problem to leave the safety on, and highly unrealistic. In terms of realism of using a weapon (i.e. would the weapon do this in real life?) it is a catastrophe.
    - This may be related to finger fatigue. It requires muscle control to hold your finger above the mouse button, and as the muscles fatigue, the finger drops.
    - Concur, but marginal impact on desktop set
    - Sometimes with the M4/203 it's easy to hit the right mouse click and toss a grenade.
    - If it is really necessary a separate "safety" key or user call out can be implemented into the training routine.
- Recommendation
    - Create a Safety control on the desktop system that prevents both primary and secondary fire (recommend Caps Lock).
    - Either less sensitive buttons on the mouse (best solution), or possibly train users to hold the mouse differently (not best solution)
    - If it's really necessary implement a procedure or key to put on a safety. I know for a fact the Source engine has a "holster weapon" key, where you can put down your active weapon.

Concern D03

Description. Accidentally pressing the Windows key (between Alt and Control) switches to the desktop and opens the start menu. You can just switch back to the simulation, but sometimes it will not load properly and requires a full restart.

Table A-15. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern D03

	Frequency	Impact	Persistence
Mean	1.25	2.625	2.125
Mode	1	3	3
SD	0.886405	0.744024	0.834523

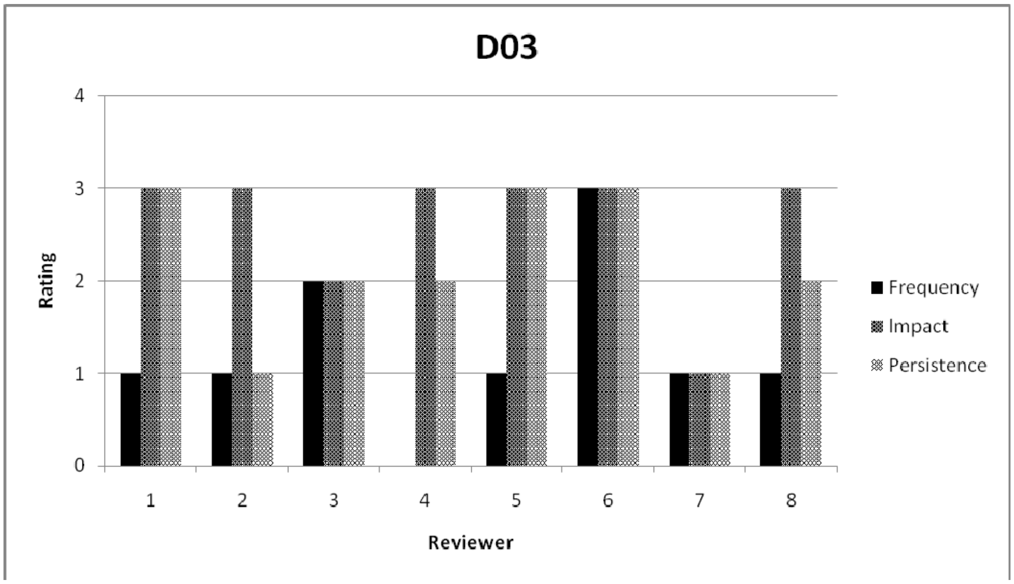

Figure A-15. Evaluators' Ratings for Concern D03

Evaluators Comments

- General comments
  - This never happened to me, nor have I seen it happen to anyone else. It is probably one of those things that is extremely difficult to fix.
  - I did not experience this problem, but as it relies mostly on the user hitting the key by accident (which hopefully will not occur too often) and then "sometimes" not reloading properly, I would not make this priority.
  - Never had this occur. But if it did it could obviously be a big problem even if it happens one time. Not sure how to answer the persistence question?
  - This is one time when you cannot undo a mistake. Also, the windows key is very close to the shift and control keys which are used frequently in the simulation and therefore there is a greater chance to accidentally hit the windows key.
  - concur - need to disable this function
  - I didn't experience this, but the windows key can be turned off.
- Recommendation
  - This may be unavoidable on Windows systems. Disable the Windows key's functionality within Windows if possible, or use keyboards without Windows keys if they are available. If neither of these are feasible options, instruct users of this issue so that they know to try to avoid it.
  - Another non-issue.
  - This can be avoided by disabling the Windows key's functionality by editing registry files.

Concern W08

Description. The simulated weapon in the wearable system works differently than real weapons (unable to switch between Semi/Auto, charging handle isn't used when reloading, no separate trigger for firing rifle grenades).

Table A-16. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern W08

	Frequency	Impact	Persistence
Mean	2.25	1.875	1.875
Mode	2	2	2
SD	1.164965	1.246423	1.246423

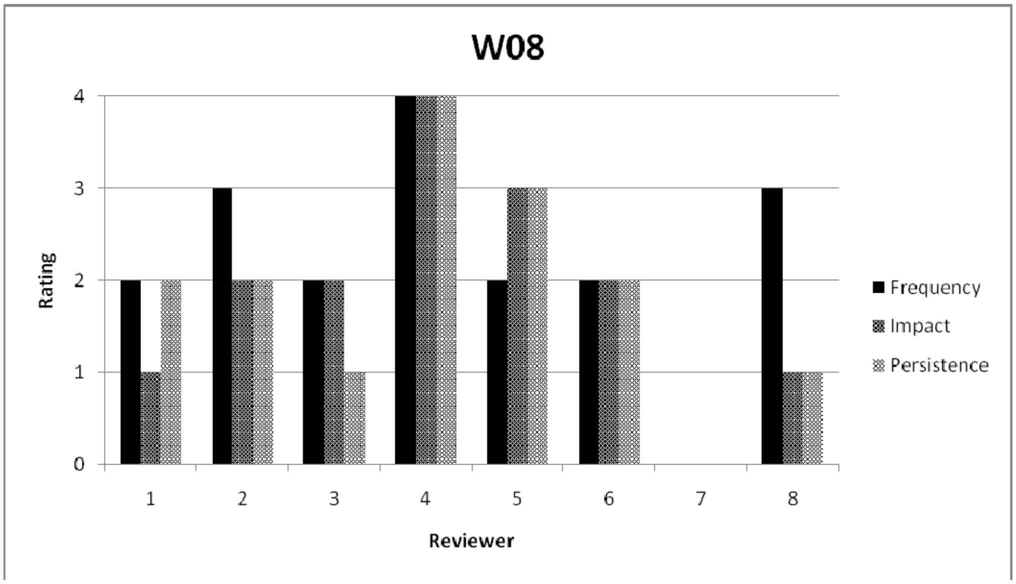

Figure A-16. Evaluators' Ratings for Concern W08

Evaluators Comments

- General comments
  - There is the possibility of negative transfer of training for people trained to use real weapons (i.e. Soldiers)
  - yes, concur
  - This can be changed with a new weapon. Again as this is a training tool some concessions must be made.
- Recommendation
  - Create functional fire select lever, charging handle, and secondary fire trigger.
  - Not a real concern.

Concern B01

Description. No Help system is provided, though it is arguable whether one should exist outside of providing information for the controls

Table A-17. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern B01

	Frequency	Impact	Persistence
Mean	2	2	1.625
Mode	2	2	2
SD	1.069045	0.534522	0.517549

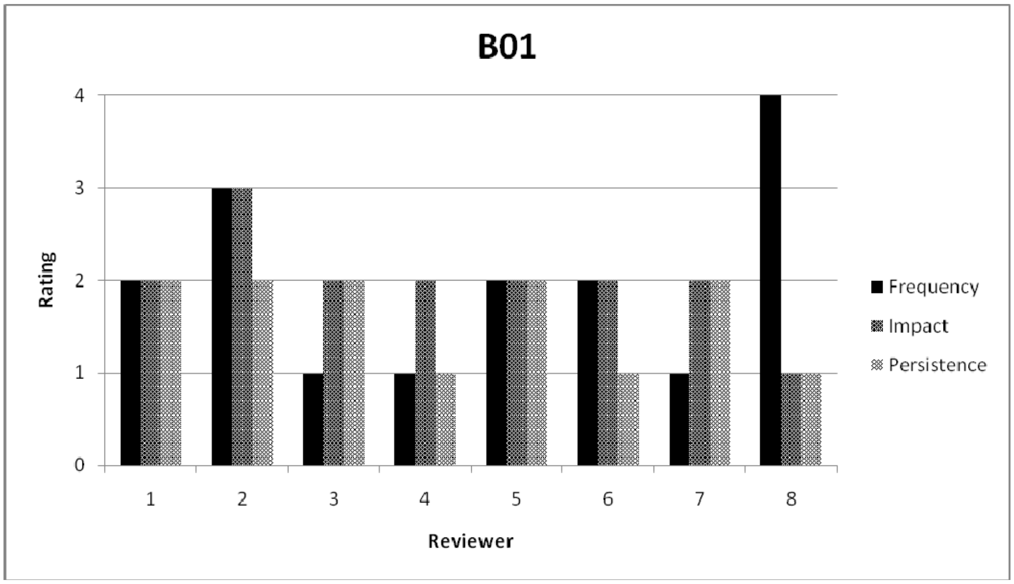

Figure A-17. Evaluators' Ratings for Concern B01

Evaluators Comments

- General Comments
  - If the user receives enough training to learn the control combinations, this problem will go away.
  - With multiple users operating as a team, a single user would not have time to access help without slowing down the team.
  - In the style of the old usability adage: if it needs instructions it's not a good design.
- Recommendation
  - Ensure that instructors are available to provide initial instruction for new users, and resolve any issues encountered by users throughout training. This will work better than any training manual that could be provided, and although costly, seems to be necessary for this type of system.
  - Is a help system really needed? A Soldier in combat would fall back on training expertise.
  - Controls should be somewhat intuitive, perhaps users should be able to customize their controls and the experimenters can see if a trend develops.

Concern W01

Description. Modifying the controls can't be done in the system itself, it requires connecting to an external keyboard/monitor and editing a text file. Not all button combinations are available (i.e. can't assign secondary fire to 1+Trigger).

Table A-18. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern W01

	Frequency	Impact	Persistence
Mean	1.875	1.5	1.857143
Mode	1	1	2
SD	0.834523	0.547723	0.690066

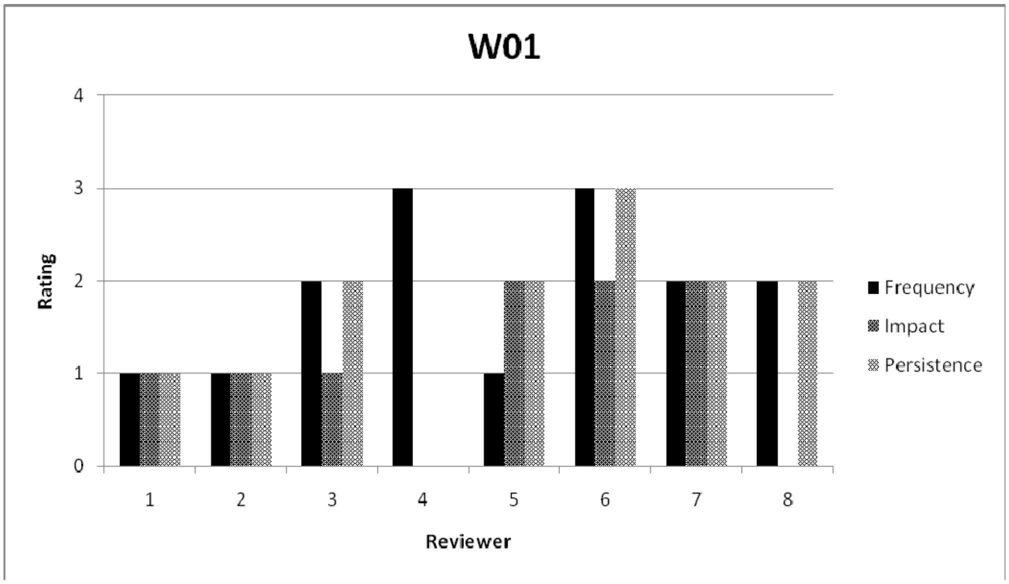

Figure A-18. Evaluators' Ratings for Concern W01

Evaluators Comments

- General comments
    - Realistically, we are not going to want to change the controls that frequently, although this comment does raise the general problem of hard to remember and hard to use controls.
    - Any gamer will want to set up the controls how they want them and how they are familiar with them because they set them up that way in every FPS they play. I would assume a gamer would change the controls if the GDIS controls were not in line with contemporary gaming standards. It was my experience that most of the controls were manageable. A user should be able to adapt their play to the controls in relatively short order. The ability to remap the keys would be a benefit, but in my opinion, would be a nice treat to the user, rather than a necessity.
    - Obviously it could be a problem for flexibility and realism...people don't change the controls on a real gun on the fly. On the other hand, for game play, users can probably learn the conventions for the controls however they are set. But transfer to a real world task would make this a catastrophe.
    - Generally concur.
    - Agreed.
- Recommendation
    - While it may not be feasible to modify controls from directly within the system, it should at least be done within the software once an external keyboard is connected, rather than having to completely exit the simulator to modify a text file. All system controls (and all combinations of system controls) should be capable of being mapped to any system function.
    - A lot more practice time before an exercise is conducted
    - Work with the tool developers to add some freedom to the system.

Concern B06

Description. Mistakes can be easy to make by accidentally pressing a button (note: this does not include accidentally firing the weapon, which is listed separately).

Table A-19. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern B06

	Frequency	Impact	Persistence

Mean	1.75	1.375	1.875
Mode	1	1	2
SD	1.28174	1.30247	1.246423

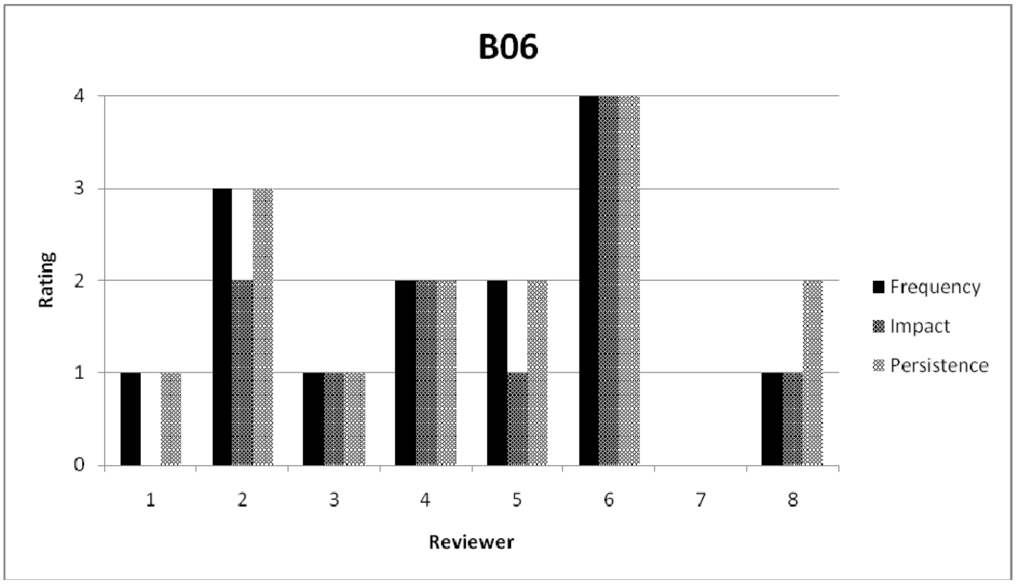

Figure A-19. Evaluators' Ratings for Concern B06

Evaluators Comments

- General comments
  - Recover is usually easy,
  - There may be negative transfer problems with users who play other games with different controls.
  - The wearable set was where I made my mistakes by pressing buttons...
  - Not so much in the wearable.
  - Training is there to reduce the frequency of mistakes
- Recommendation
  - None. Mistakes are generally not disruptive to the task and are easy to recover from. Learning to recover from one's mistakes is an important part of training, and so the simulator should allow them to happen.
  - Limit control inputs
  - The gun on the wearable system is kind of dumb with the four buttons and very unforgiving simultaneous pressing system. This should be altered to make the buttons perhaps have fewer functions requiring less precise timing or a "shift key" that you push first to get the other functions

Concern B11

Description. Direction of movement in both systems is directly tied to the direction you are looking.

Table A-20. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern B11

	Frequency	Impact	Persistence
Mean	1.857143	1.428571	1.142857
Mode	3	0	1
SD	1.573592	1.397276	1.069045

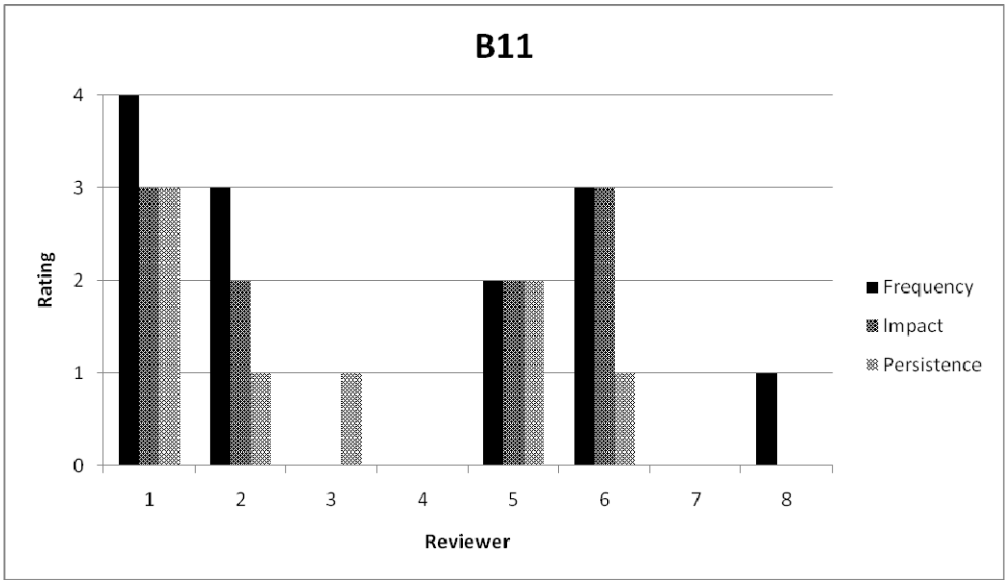

Figure A-20. Evaluators' Ratings for Concern B11

Evaluators Comments

- General comments
  - There are ways to separate direction of view and direction of movement, but it would add to the complexity of the controls in the desktop, and would introduce another source of error in the wearables.
  - So is the flashlight, which is unrealistic since the flashlight is mounted on the weapon. This may cause negative transfer of training.
  - This is false. You can use the strafe keys
- Recommendation
  - This seems to be a fundamental problem with desktop simulators, with no apparent solution. For the wearable system, direction of movement should be based on a separate sensor, such as the thigh tracker. However, with the current design the thigh tracker is prone to sliding around on the leg throughout the exercise, which will likely hinder movement more so than the current system. The thigh tracker must be made more secure, or an additional tracker may be needed on the chest or belt specifically for determining direction of movement.
  - Do not change.
  - Needs work in both apps.
  - People should get better at playing videogames. In all seriousness a longer familiarization session may be implemented so folks "get the hang of it."

Concern D02

Description. The printed control sheet is needed for remembering the controls with the desktop simulator.

Table A-21. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern D02

	Frequency	Impact	Persistence
Mean	1.571429	1	1.428571
Mode	3	0	2
SD	1.272418	1	1.133893

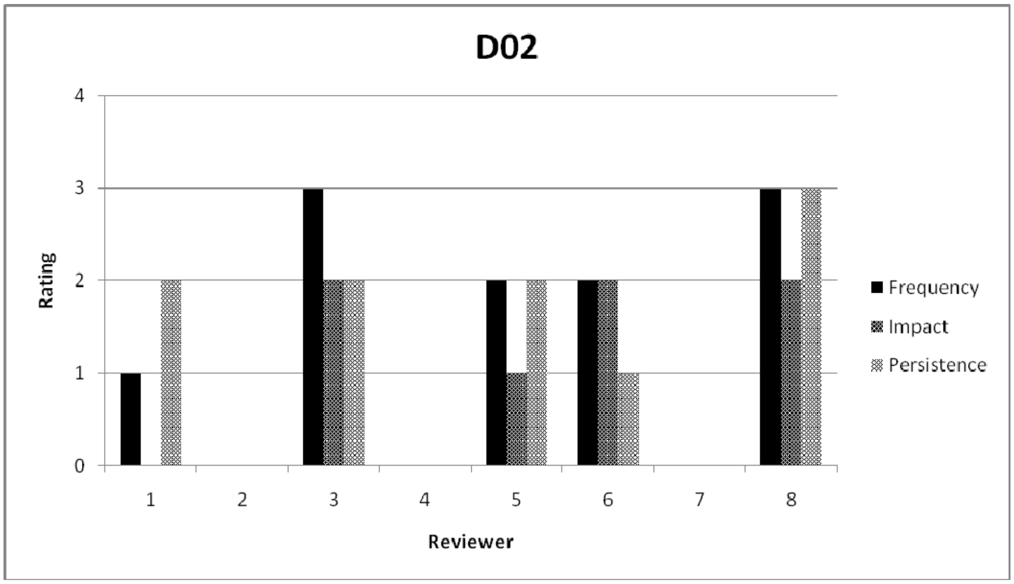

Figure A-21. Evaluators' Ratings for Concern D02

Evaluators Comments

- General comments
  - This really is not a problem, is it?
  - The printed control sheet is helpful, if it were not there it would be a minor problem, and would go away with training.
  - The control sheet is very helpful for persons who are not familiar with the control conventions.
  - It was fine
  - The sheet is helpful.
- Recommendation
  - Researchers/Trainers using this, or similar systems, should provide similar reference sheets to their participants/trainees. Or: The software should provide a quick control reference within the system itself, that adapts to any modified control mappings, and organizes commands in a logical structure so that they can be found quickly during an exercise. It should also be capable of displaying only the controls to be used in a given exercise, as defined before hand by a system administrator.
  - Continue to use the printed sheet.
  - Perhaps something similar may be devised for the wearable element.

Concern B03

Description. The amount of weapons/ammunition that can be carried is potentially unrealistic.

Table A-22. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern B03

	Frequency	Impact	Persistence
Mean	1.25	1.125	1.375
Mode	1	1	1
SD	1.28174	1.356203	1.30247

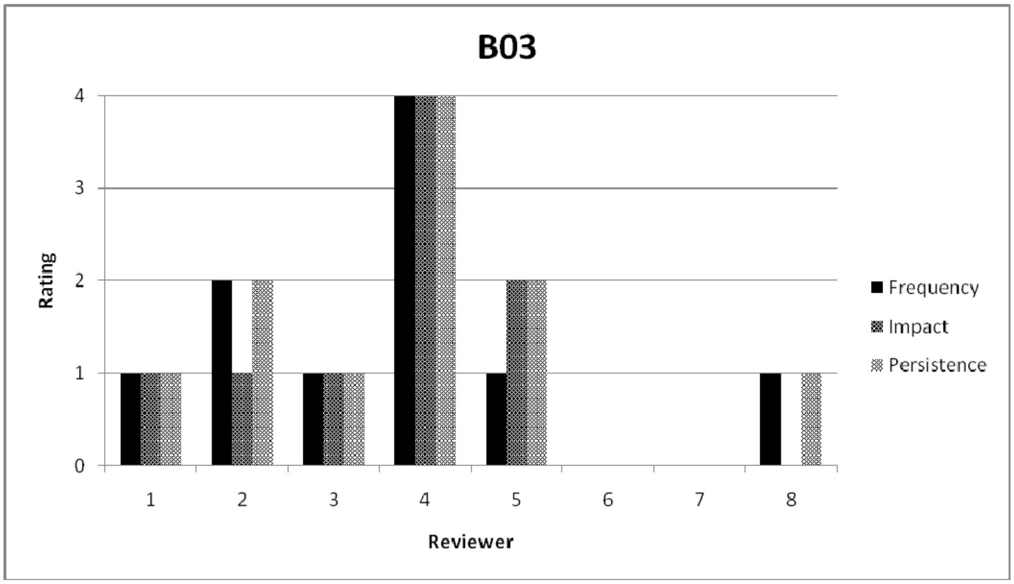

Figure A-22. Evaluators' Ratings for Concern B03

Evaluators Comments

- General comments
  - This is more of a scenario issue than a system issue. I think it should be a quick and easy fix.
  - I don't recall an unrealistic basic load or infinite reloads.
  - This is nonsensical in terms of usability, this can be easily altered.
- Recommendation
  - The weapons/ammunition avatars load with is easily modifiable by system administrators; they need to keep realism in mind when setting these parameters.
  - Fix in scenarios
  - Avatars should be able to carry a realistic basic load of ammunition, and should be able to restock the ammunition load to simulate resupply. The weapons load should be restricted to simulate what a Soldier would carry, e.g., one shoulder weapon and one pistol.
  - Model the ammo count after what your standard Soldier carries.

Concern B10

Description. The researchers/instructors provided necessary help/instruction throughout the session.

Table A-23. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern B10

	Frequency	Impact	Persistence
Mean	1.375	1	1.142857
Mode	0	0	2
SD	1.59799	1.154701	1.069045

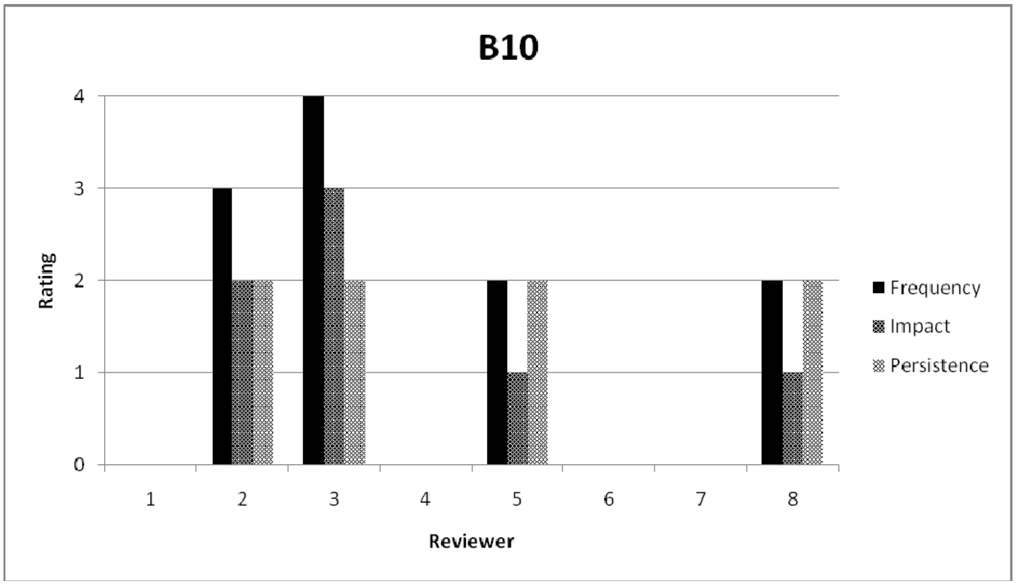

Figure A-23. Evaluators' Ratings for Concern B10

Evaluators Comments

- General comments
  - This was an artificial situation. However, this comment does illustrate the need for more training than was provided.
  - Help/instruction was frequent, the impact of it NOT being there would have been major and if it had not been there, with training it may not matter.
  - The instruction was helpful, but learning and remembering how the weapon worked was difficult.
  - Overall good instruction and real-time help.
  - Researchers provided help when asked.
- Recommendation
  - Always ensure that there are instructors/researchers/system operators available to solve problems and answer questions as they arise. The number of instructors needed will vary based on the number of trainees and their level of experience with the system.
  - Put them in the simulation as a virtual team-mate/helper
  - This was ok

Concern D01

Description. Some of the desktop controls do not work as initially expected (example: "O" for map/compass).

Table A-24. Mean, Mode, and Standard Deviations of Frequency, Impact, and Persistence Ratings for Concern D01

	Frequency	Impact	Persistence
Mean	1.375	0.75	1.125
Mode	1	1	1
SD	1.30247	0.707107	0.834523

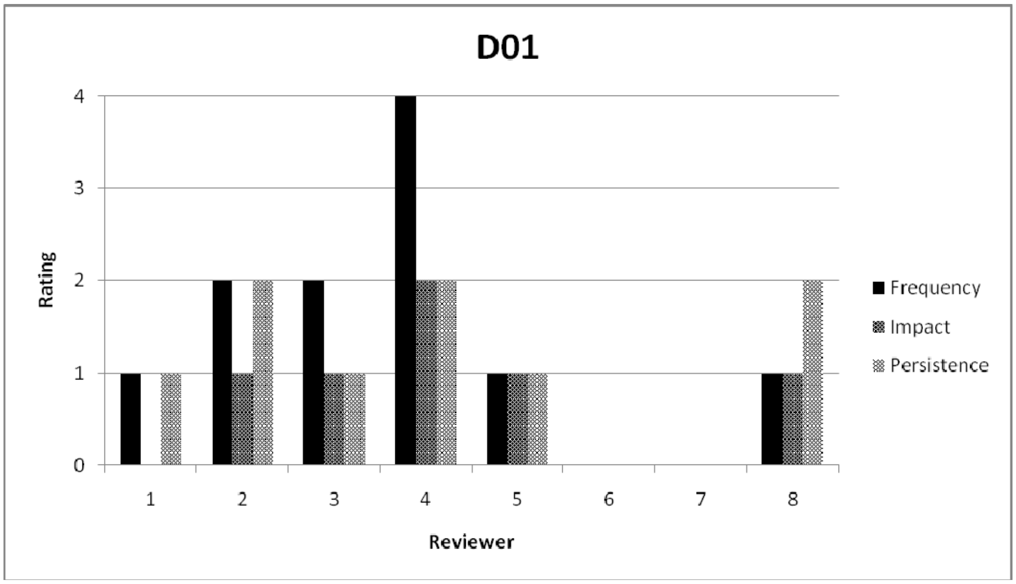

Figure A-24. Evaluators' Ratings for Concern D01

Evaluators Comments

- General comments
  - After or one or two mistakes the user should be familiar enough not to make the mistake again.
  - No issues
  - These can always be altered. Users may customize controls in a way that makes sense to them.
- Recommendation
  - None, the controls are generally learned relatively easily with time in the simulator. The system is already capable of remapping the controls to other keys if this is deemed necessary.
  - I think it should be fairly easy to change the control keys if necessary.
  - Worked as advertised
  - Look for trends if people prefer to alter control schemes.

Heuristic Evaluation Worksheet

Simulator: Desktop / Wearable		Evaluator #:

Instructions: For each of the heuristic principles, rate the simulator on the Likert scale, and then list any usability concerns you encountered. You can make any general comments in the space provided. The back page has space for notes or additional comments.

1. Visibility of system status
The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

"It's easy to determine the system status and know what's going on" (circle one):

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Not Applicable

Usability Concerns:

General Comments:

2. Match between system and the real world
The system should speak the users' language, follow real-world conventions, and make information appear in a natural and logical order.

"Information is easy to understand and is presented in a natural and logical order" (circle one):

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Not Applicable

Usability Concerns:

General Comments:

3. User control and freedom

Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

“If I make a mistake it’s easy to undo and then choose the correct function” (circle one):

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Not Applicable

Usability Concerns:

General Comments:

4. Consistency and standards

Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

“All functions follow consistent system conventions” (circle one):

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Not Applicable

Usability Concerns:

General Comments:

5. Error prevention

Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.

"The system helps me avoid committing errors" (circle one):

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Not Applicable

Usability Concerns:

General Comments:

6. Recognition rather than recall

Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

"Information I need is easy to find" (circle one):

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Not Applicable

Usability Concerns:

General Comments:

7. Flexibility and efficiency of use

Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

"There are shortcuts that speed things up" (circle one):

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Not Applicable

Usability Concerns:

General Comments:

8. Aesthetic and minimalist design

Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

"There was no extraneous or unnecessary information or objects" (circle one):

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Not Applicable

Usability Concerns:

General Comments:

9. Help users recognize, diagnose, and recover from errors *Error messages should be expressed in plain language, precisely indicate the problem, and constructively suggest a solution.*
"The system helps me recover from errors" (circle one): Strongly Disagree / Disagree / Neither agree nor disagree / Agree / Strongly Agree / Not Applicable
Usability Concerns:
General Comments:

10. Help and documentation *Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.*
"When I needed help, the help function was truly helpful" (circle one): Strongly Disagree / Disagree / Neither agree nor disagree / Agree / Strongly Agree / Not Applicable
Usability Concerns:
General Comments:

STOP! Continue to the next phase of the evaluation.

NOTES/ADDITIONAL COMMENTS