Usability Analysis of VR Simulation Software

Anna L. Weaver, Paul N. Kizakevich
RTI, 3040 Cornwallis Road, Research Triangle Park, NC 27709

Walt Stoy
Center for Emergency Medicine, 230 McKee Place, Pittsburgh, PA 15213

J. Harvey Magee
U.S. Army Medical Research and Materiel Command, Ft. Detrick, MD

William Ott
CPCS Technology, Cary, NC

Kevin Wilson
Durham County EMS, 3643 N. Roxboro Road, Durham, NC 27704

1. Background/Problem:

Usability testing, commonly conducted for commercial software to ensure that it meets the needs of the end user, is likewise vital to creating effective training software employing VR technologies. However, developers often limit software testing to expert evaluation (by other developers or subject matter experts [SMEs]) to identify potential problems that might impede the usability or acceptability of the software. This paper describes a usability test of VirtualEMS™, a VR software package designed to provide realistic practice for emergency medical technicians (EMTs) and paramedics, thereby providing insight into the process and value added by the usability testing.

2. Test Design

Following Rubin's *Handbook of Usability Testing* [1], iterative usability testing of VirtualEMS was conducted with representatives of the end user: students and teachers in emergency medical services (EMS) programs and practicing EMTs and paramedics (referred to as firehouse users). Students and teachers were tested at the Center for Emergency Medicine of Western Pennsylvania (CEM) in Pittsburgh, PA, in mid-February 2001; testing with firehouse users took place at the Durham EMS Base in Durham, NC, in late March 2001. Following CEM testing, a set of preliminary recommendations were provided to developers, many of which were implemented prior to firehouse testing.

CEM and firehouse testing provided quantitative and qualitative information about software usability, including adequacy of the software's tutorial and help system, and provided preference information (e.g., Do these users enjoy working with VirtualEMS? Do they think it provides meaningful training and/or practice? Would they be likely to use the software?).

Consistent with prior SME evaluations conducted on this software and concurrent with CEM testing, experts in emergency medicine evaluated the accuracy and realism of several aspects of VirtualEMS, including organization and completeness of tools and

devices; accuracy and realism of the visual representations of the patient, his injuries, and medical devices; and accuracy and realism of patient interactions and physiological responses represented in the software.

3. Method

Standard usability testing methods [1] employed included scripted scenarios, pre- and post-test questionnaires, data logs, the think-aloud protocol, and test monitor observations. The test monitor measured and recorded certain aspects of test participant performance, including observations and comments (test participant exhibits frustration or satisfaction, or comments about the software positively or negatively); number of noncritical errors (an individual participant makes a mistake but is able to recover and complete the task); number of critical errors (an individual participant makes a mistake and is not able to recover and complete the task); time to complete tasks (when applicable); and number of times users accessed help screens and/or the user's guide. It is important to note that the test monitor did not offer participants any assistance with the software or assigned tasks.

SMEs were asked to self-record their evaluations, although an observer was occasionally present. In contrast to the test monitor role described above, the SME observer interacted freely with the SMEs, often providing explanations and assistance with software features.

4. Usability Results

Usability testing identified problems with and potential improvements to the tutorial, help system, and other elements of the interface (mouse commands, pop-up menus). Test participants also provided an overall rating for the software.

Tutorial

Perhaps the most important qualitative finding related to user immersion in the scenario and the importance that the software tutorial follow standard EMS procedures. Several users became frustrated when the tutorial, which was designed only to teach users to operate the software, not fully follow emergency procedures. Both CEM and firehouse testing revealed that users are so quickly immersed in the simulation, that they have a strong desire to treat the patient correctly. Hence, some were so distracted by this conflict that they strayed from the tutorial. These users took far longer to complete the tutorial and occasionally became visibly agitated. Although an encouraging indication that VirtualEMS engages its target users, it often conflicted with users' ability to learn from the tutorial.

Help System

The initial Help System was web based, with links from software elements to HTML pages offering helpful information. All users consulted help files when they could not figure out a software feature or became confused. If they did not find the answer immediately, they expected to be able to search the help files and indicated that searchable help was key to overall usability and user-friendliness. Although web-based help seemed like good idea, users offered that a Windows®-compatible help system would be preferred.

Interface

Users indicated that they expect familiar Windows® interface features (e.g., cursor displays as an hourglass when the program is busy). Problems and inconsistencies in the

interface identified during testing included menus closing automatically before novice users were able to read them and message boxes sometimes obscuring text.

In addition, both CEM and firehouse users demonstrated significant problems with the "learn mode" (designed to provide users with a review of EMS assessment protocols), the software's simulation of scene assessment, and one of the primary navigational tools.

Overall Rating

On a rating scale from 1 (lowest or poorest) to 5 (highest or best), the average overall ratings given by students and firehouse users are indicated in Table 1.

Table 1. Average Evaluation Scores — Students and Firehouse Users

Test User	Overall usefulness	Meaningfulness for training/ practice	Likelihood of using outside classroom/work environment	Likelihood of using if approved for continuing ed.
EMS Students	3.4	4.0	3.7	**
Firehouse Users	4.0	4.8	4.8	4.3

** Participant was not asked this question.

5. SME Results

SMEs identified several problems with accuracy and realism of the software. For example, visual representations of some wounds were determined to be unrealistic, and vital signs were determined to be "way too good" for the severity of injuries depicted in most scenarios. Accuracy, realism, and function of patient interactions and medical devices were deemed adequate for EMT training.

6. Conclusions

Usability testing identified markedly different problems than SME evaluations, thus proving that usability testing does add value to the software development effort. Problems with the interface that might have gone unnoticed in SME evaluations were identified and recorded, and their effect on overall user satisfaction was realized. Observing actual end users working with software can be key to developing training software that will be accepted by users and will, thus, achieve the intended training goal.

After the first round of tests (CEM), several changes were implemented. As shown in Table 1, subsequent firehouse testing revealed improvement in users' perception of the software's meaningfulness and overall usefulness, and an increase in the likelihood that users would use the software. (Because of the small sample size, formal hypothesis testing was not done). This would indicate that the changes recommended after initial usability testing increased user satisfaction with the software.

The usability testing methods applied in this study proved adequate for testing desktop-VR software like VirtualEMS™; however, other accepted usability evaluation methods might also be applicable, including heuristic evaluation (usability specialists judge whether each element of a user interface follows established usability principles), cognitive walkthrough (expert evaluators construct task scenarios from a specification or early prototype and then role play the part of a user working with that interface), and pluralistic walkthroughs (users, developers, and usability professionals step through a task scenario, discussing and evaluating each element of interaction) [2].

7. References

[1] Rubin, J. (1994). *Handbook of usability testing*. New York: John Wiley & Sons.

[2] Hom, J. (1996). *The Usability Methods Toolbox* [Online]. Available: http://www.best.com/~jthom/usability/usable.htm

This software was supported in part by the U.S. Army Medical Research and Materiel Command under Contract No. DAMD17-99-2-9046. The contents of this software does not necessarily reflect the position or the policy of the government, and no official endorsement should be inferred.