Virtual Reality for Training: Evaluating Knowledge Retention

Craig R. Hall
Air Force Research Laboratory .
Brooks AFB, TX

Randy J. Stiles
Lockheed Martin Advanced Technology Center
Palo Alto, CA

Carol D. Horwitz
Command Technologies, Inc.
Brooks AFB, TX

Abstract

Skill decay after periods of skill disuse is well known and has substantial implications when relatively long periods of time separate training from the application of learned skills. We conducted a small study that examined the differential effects of virtual reality versus conventional computer-based media on skill retention. The results reported are preliminary, but were consistent with earlier research that reports that VR may not be superior to conventional electronic media for training certain intellectual skills. Little is known, however about the effects of VR in support of practice strategies for reducing skill decay. Implications for future research are discussed.

Introduction

Skill retention is important because of the time that potentially separates skill acquisition and use of the same skills on the job. There are many situations that demand the application of skills that have not been applied for extended periods of time. Despite the apparent importance of skill retention on performance, however, few empirical studies have tested factors that mediate skill retention [3]. Some factors have been associated with increasing skill retention, such as extended practice of skills. While practice has been established as a strong predictor of skilled performance, it has not been established whether authentic contexts, such as those found on the job, are related to retention. Further, the cost of simulating authentic contexts in the laboratory has traditionally been prohibitive. The rise of virtual reality (VR) technology promises the creation of low-cost, authentic job contexts in which the full range of intellectual and psychomotor skills can be acquired and practiced. Despite VR's appeal, however, a framework for guiding the selection and design of effective VR-based training has not been developed and much work remains to be done before the training potential of VR is fully known. For example, little is known about how, or if, VR is superior to conventional computer-based media for enabling skill development. Investigation into learning effects from media have a long history and there may be evidence that learning is more related to the instructional strategies employed with the media, rather than from the attributes of the media [1]. While most media comparisons have examined the differential effects of media on achievement, little is known about how media attributes facilitate skill retention after training.

An experiment on knowledge retention

This paper reports the outcome from a pilot study that addressed the effects of alternate presentation modes on knowledge retention. The primary purpose of this study was to aid in the structuring of a more substantial study to be conducted in October 1997. As a first step in our research, we have chosen to establish the instructional utility of VR-based instruction as it relates to retention of knowledge. This study was conducted by the Air Force Research Laboratory, Brooks Air Force Base, San Antonio, TX, as part of the Virtual Environments for Training (VET) program that is being funded by the Office of Naval Research (Contract N00014-95-C-0179). It is a Defense Department-focused research initiative to address technical issues in applying virtual environments to training.

One presentation mode engaged learners in the use of computer-based multi-media in which information was displayed using two-dimensional images on a conventional computer monitor (2D group). The other mode engaged learners in the use of virtual-reality-based media which allowed the learner six degrees-of-freedom movement in a life-size virtual room with 3D models of the devices used in the procedure (VR group). We expected that subjects who learned the procedure via navigation in the virtual space, allowing interaction with life-size virtual equipment, would incorporate their spatial experiences with their learning of the procedure, resulting in higher recall of the procedure over time. The learning goal of the courseware was recall of a ten-step procedure for operating several devices distributed around a room. The devices comprised a large air compressor used by the Navy on some of its ships.

Subjects were ten civilian and military personnel from Air Force bases located in and around San Antonio. There were three males and two females in the VR group and four males and one female in the 2D group. Ages of subjects ranged from late-twenties to early-fifties. The research was conducted on a Silicon Graphics computer running the Unix operating system. The virtual environment was presented in a fully immersive interface. This was accomplished using PINCH gloves from Fakespace, Virtual Research's V6 head-mounted display (HMD), and three Flock of Birds trackers from Ascension Technology. The virtual environment was rendered with Vista Viewer, a Silicon Graphics/Performer-based software agent that provides an advanced 3D interactive display [4,8]. The 2D human-computer interface, on the other hand, consisted of a conventional windows-based display and computer mouse. The instructional software for both the 2D and 3D conditions was comprised of lessons in the VIVIDS© Authoring System, designed to cost-effectively develop, deliver and maintain simulation-based tutors for field and laboratory applications. VIVIDS is being developed by Behavioral Technology Laboratories of the University of Southern California under contract to Air Force Armstrong Laboratory, and is based on its predecessor, RIDES© [5,6]. Instruction for both the 2D and VR groups was delivered in the form of a verbal narrative using Trish, a text-to-speech software agent based on Entropic's Truetalk.

In the 2D group, subjects used software that provided flat representations of the devices used in the to-be-learned procedure. The software displayed a floor plan of the room and the equipment used in the procedure. Selecting a machine or device in the floor plan with a computer mouse opened a separate window showing a photograph or illustration of the object. Subjects used the computer mouse to interact with buttons and other features of the objects in their windows. The instruction prompted the subject to open object windows in the sequence of the procedure.

In the VR group, subjects used software, which displayed the virtual room and three-dimensional representations of the devices on an HMD worn by the subject. In this fully immersive interface, the subject could look around the virtual environment by physically turning his or her head. Movement about the room was accomplished via the PINCH gloves worn by the subject, as was manipulation of device controls. For example, pressing the middle finger and thumb of the right hand simulated forward movement. Pressing the forefinger and thumb together while intersecting a control on a virtual device constituted manipulation of that control. The instruction prompted the subject to move to each device in the procedure and manipulate it.

The instructional content for both groups was identical, except where navigational differences in the human-computer interface required special instructions. For example, 2D subjects were prompted to summon representations of equipment by clicking the mouse-pointer on icons in a floor plan of the equipment room. In contrast, VR subjects were merely instructed to move to a device in the equipment room.

The instructional strategy provided by the software consisted of verbally prompting the subject to locate devices in the procedure and to provide remediation when the subject selected devices not included in the procedure or in the wrong sequence. There were two phases to the instructional strategy. The first phase instructed subjects to select each device in the procedure. Each time the subject correctly selected a device, the narrative informed the subject of the name and location of the next device. Devices were not labeled. To enable rapid identification, devices were highlighted by alternating their colors in the manner of a flashing beacon. Flashing terminated when the subject selected a device. Following the first phase of instruction, the second phase prompted subjects to practice the procedure. The verbal instructions consisted of informing the learner which device was next in the procedure, but did not inform the learner about the location of a device. A device did not flash until the learner made two incorrect selections or selected a "don't know" button, at which time, verbal instruction about the location of the device was supplied. Subjects were free to repeat the procedure until they reached mastery or exceeded the time allotted, which was 45 minutes. Each time a subject completed the procedure, the computer informed him or her of the number of devices correctly selected (e.g. "...your score was 8 out of 10").

Instructional time and number of practice trials were allowed to vary freely (up to 45 minutes) because we wanted to determine the range of variability for those factors. We suspected that the facility with which subjects manipulated the human-computer-interfaces in each treatment was a primary source for the variance in instructional time and total practice trials. Subjects in the VR condition were expected to simulate walking between goals in the procedure, which would take substantially more time than it would take to move a computer mouse between goals in the 2D condition.

Achievement was tested using a scale model of the devices included in the instruction. The achievement scores were based on (1) the number of devices recalled in the correct sequence and (2) the number of device names correctly recalled. A proctor recorded the subjects' performance on a floor plan of the scale model while observing subjects execute the procedure. There was only one correct path for executing the procedure. We predicted that subjects in the VR group would score higher on both measures of achievement than those in the 2D group.

Each subject followed the same general plan: (a) orientation to the task, (b) interaction with the courseware, (c) a five-minute post-test immediately after, followed by (d) a five-minute post-test seven days later. Orientation to the task consisted of informing the subject about the nature of the procedure to be learned. Orientation procedures for the VR group included an additional task, in which subjects practiced navigating around a simple environment so that they could master manipulation of the interface. Subjects practiced operating the virtual equipment until they felt confident in their ability to maneuver and manipulate device controls in the virtual environment. 2D subjects did not practice with the 2D interface because all subjects were daily users of Windows operating systems. During interaction with the courseware, subjects relied on the automated instruction to learn the procedure. A proctor recorded all problems that subjects experienced and any behaviors those subjects manifested while interacting with the hardware and software.

Qualitative analyses were used in this study. The small sample size (n = 10), coupled with the exploratory nature of the study, precluded reliable use of statistical tests. The data, however, were examined for trends. Visual inspection was performed to determine whether relationships were suggested for (a) treatment and achievement, (b) treatment and knowledge retention, and (c) level of manipulation facility and achievement. Other observations aimed at determining the nature of subjects' interaction with the software were also analyzed for indicators related to achievement.

Results

A procedural test was scored by summing the correct number of steps that subjects pointed to in the correct sequence on the scale-model of the air compressor room. The test was administered immediately after the instruction, and again one week later. Table 1 reports the means and standard deviations for both tests.

Table 1 - Procedural Test

	Immediate		One Week Later	
	M	**SD**	**M**	**SD**
2D	**7.1**	**1.7**	**6.5**	**1.4**
VR	**8.4**	**1.3**	**6.9**	**1.7**

Note. Tests were graded on a 10-point scale.

The VR group generally scored higher on the immediate procedure test than did the 2D group. In contrast, performance one week later appeared to be nearly the same across both groups. The amount of knowledge loss between the immediate and delayed tests was greater for the VR group than for the 2D group.

Summing the correct number of devices that subjects correctly named in the procedure scored a test of recall of names. The name test was administered concurrently with the procedure test. Table 2 reports the means and standard deviations for both name tests.

Table 2 - Name Test

	Immediate		One Week Later	
	M	**SD**	**M**	**SD**
2D	**5.7**	**1.7**	**5.6**	**1.9**
VR	**7.0**	**1.3**	**6.8**	**0.7**

Note. Tests were graded on a 10-point scale.

The VR group generally scored higher on both the immediate and delayed name tests than did the 2D subjects. The amount of knowledge loss between the immediate and delayed name tests was small and nearly the same for both groups.

Analyses examined the relationship between manipulation facility and achievement. The level of manipulation difficulty was determined by observing subjects as they worked their way through the instruction. The VR subjects were observed putting much more effort into using the interface than did 2D subjects. All VR subjects initially experienced difficulty with using the VR interface, but improved their facility as they worked their way through the instruction. Two VR subjects, however, tended to have difficulty maneuvering through the environment and reported the most physical discomfort, complaining of slight nausea. Some VR subjects complained of heat buildup under the HMD. Most VR subjects were visibly sweating around the temples and brows upon removal of the headgear.

VR subjects spent considerable time navigating between points in the environment, and spent much more time than did 2D subjects trying to operate virtual devices. VR subjects required an average of 9 minutes to familiarize themselves with operating the VR gear. 2D subjects spent almost half as much time with the instruction (21 minutes) as did VR subjects (38 minutes) and 2D subjects generally practiced the procedure more times (3.3) than did the VR subjects (2 times). 2D subjects were able, within the allotted time of 45 minutes, to repeat the procedure until they reached mastery. In contrast, VR subjects either ran out of time or became too exhausted before they could reach mastery.

Observations also included how subjects interacted with the instructional strategy. Subjects generally followed instructions and appeared to follow the instructional strategy in the same manner. However, one VR subject and one 2D subject were observed to impose their own strategies for learning the procedure. Each time these two subjects progressed to the next step in the procedure, they paused to rehearse the entire procedure from the beginning by pointing to each device in the proper sequence. These two subjects' average scores were the highest in their respective groups. Other VR subjects were observed to vocalize some of the goals as they navigated to them, especially during extended searches for devices. In several instances, VR subjects asked for the last instruction to be restated. Restatements were requested after subjects had effected extended searches for a device, and when the subject reported being unable to understand the computer-generated voice.

Discussion

While the trends in the test-scores appeared to favor VR, the small sample size of this study precluded the formation of reliable inferences. However, even if the trends found here could be replicated by using a larger sample, it is unlikely that the media attributes in this study account for learning differences. In this study, the basic instructional strategy was the same, but the amount of effort that students invested while navigating between goals was different. Students in the VR group appeared to invest much more effort while navigating between steps in the procedure than did 2D subjects. Learning was probably greatest when subjects had the opportunity to actively rehearse the steps in the procedure. Rehearsal was accomplished by a) subjects' own deliberate practice strategies and b) the rehearsal of goals while navigating to the goals. More time and effort was expended navigating between goals in the VR condition, so more time was probably spent actively maintaining the name and location of the goals in memory. Despite any apparent effort expended by VR subjects, however, scores on the delayed retention procedure test did not appear to be different between groups.

One might be tempted to explain any apparent difference in performance in this study by pointing to the greater amount of time that VR subjects spent in the instruction. However, it was observed that the 2D subjects generally practiced the whole procedure more times than did VR subjects and all 2D subjects achieved mastery before taking the tests. Most of the VR subjects, on the other hand, did not achieve mastery before taking the tests. The difference in mastery during practice may have been related to the difference in skill decay between groups on the procedure test. Skill decay for the VR group was generally greater than for the 2D group. Despite their lower initial performance on the procedure test, the 2D group may have reduced skill decay by practicing the procedure more times. Over-learning has been cited as having a negative relationship with amount of skill decay [3].

On the other hand, the difference in performance between groups appeared to be nearly the same on both name tests, while decay of name knowledge appeared to be trivial. Prior to using the courseware, all subjects were informed that they would be tested on the procedure, but were not informed that they would be tested on device names. Recall of names appeared to be lower than recall of steps in the procedure, but it was not clear why procedural knowledge decayed while name knowledge remained constant. Perhaps an interaction existed between media and knowledge type.

A question for future VR research would be: If the quantity of practice was held constant, would the quality of practice afforded by media be related to achievement? In other words, would the affordances of VR enable more effective practice than conventional computer-based instruction? In this study, the VR human-computer interface may have encouraged learners to rehearse goals. If the difficulties imposed on subjects by the VR interface actually enhanced learning, then how would learning be influenced in the VR condition if the interface was very easy to operate? If operation of the interface was completely transparent to the user, rehearsal may not have occurred, but VR subjects might have practiced the procedure more times. Subjects in the 2D condition tended to move quickly through the instruction and did not appear to rehearse the next goal before locating and selecting it, but did practice the procedure more times than did VR subjects.

Future comparisons of VR with other media should address potential interactions between interface manipulation, practice strategies and knowledge type. However, instructional courseware that fosters learning strategies should also be considered. The two highest-scoring subjects who overtly rehearsed goals appeared to employ a deliberate strategy that was independent of their experiences with either the 2D or VR interface. Although the instructional strategy of the courseware prompted subjects to practice the procedure, there were no prompts to remind subjects to employ their own intentional learning strategies. Perhaps instruction that elicits the recall or application of active learning strategies would be a useful feature for VR-based courseware. Past media studies, independent of instructional strategies, have not shown significant differences between types of media for learning skills. For example, Regian [7] conducted a study in which sixty subjects learned to operate an equipment console and to navigate around a building. One group learned using 2D computer-based instruction and another group used VR technology similar to the VR equipment used in this study. There was no statistically significant difference in performance between the groups on either of the tasks. There may be a trivial advantage afforded by authentic representations of context for the learning of relatively simple intellectual skills, such as names, route knowledge or event sequences. For more complex skills, on the other hand, there may be a substantial advantage afforded by practice strategies that are aided by authentic task representation.

Few studies have examined how media attributes can facilitate practice strategies in authentic contexts such as real work environments. Ericsson and Lehmann [2] point out that the attainment of exceptional performance is usually accompanied by sustained, deliberate practice. A central feature of deliberate practice is in the setting of performance goals and the application of practice strategies to attain goals. Learners also make use of feedback to adjust the quality of practice. If the attributes of media can be leveraged to facilitate effective practice, then there is likely to be an advantage to using, say, VR for practicing psychomotor skills, especially skills that require timing and precision that can only be acquired through deliberate practice with authentic tasks.

Finally, there is probably an economic advantage to software-based virtual environments. The rise of reusable software offers the potential to build libraries of reusable virtual devices, other simulated objects and instructional strategies. During development of the two-dimensional treatment for this study, many design decisions had to made in order to reduce the 3D nature of the content into two dimensions. In contrast, the VR world was simply built to scale. The technology employed in this study enables automatic rendering of 3D models from engineering specifications. For spatial domains, tools that enable the development of reusable instructional virtual worlds offer economic advantages that may obviate conventional two-dimensional media where high-fidelity simulations are desirable or necessary.

Summary

While there is little evidence to suggest a difference in instructional utility between VR and conventional computer-based media for some types of knowledge, there is substantial promise in the combination of VR-based practice strategies with efficient instructional development. The central focus of future research should address the facilitation of instructional and practice strategies that lead to competent application of skills in the field. We should also examine ways to support development of skills that demand the kind of activities that cannot be supported by non-VR alternatives to live contexts. We expect to find that, format limitations imposed by a conventional windows and mouse interface will impact learning of some types of tasks. For example, assembly tasks requiring timing and precision may be better accomplished when the learner has the opportunity to use hands instead of effecting actions with a computer mouse.

References

1 Clark, R.E. (1994). Media will never influence learning. Educational Technology Research and Development, 2(42), 21-29.

2 Ericsson, K.A. and Lehmann, A.C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. Annual Review of Psychology, (47), 273-305.

3 Lance, C.E, Bennett, W.R., Teachout, M.S., Harville, D.L., Welles, M.L. (in press). Moderators of skill retention interval/performance decrement relationships in eight U. S. Air Force enlisted specialties. Human Performance.

4 Horwitz, C., Fleming, J., Regian, J., and Stiles, R. (1996). Software tools for embedding principled instruction in virtual environments. In proceedings of the IMAGE Conference, Scottsdale, Arizona 23-28 June 1996.

5 Munro, A. (in press). The RIDES Simulation-based Instructional Authoring System, USAF Armstrong Laboratory Technical Report.

6 Munro, A., Johnson, M.C., Surmon, D.S., and Wogulis, J.L. (1993). Attribute-centered simulation authoring for instruction. In Proceedings of AI-ED 93 World Conference on Artificial Intelligence in Education.

7 Regian, J.W. (1997). Virtual reality for training: Evaluating transfer. In Seidel and Chatelier (Eds.), Virtual Reality, Training's Future? (pp. 31-40). Plenum Press: New York.

8 Stiles, R., McCarthy, L., & Pontecorvo, M. (1995). Training studio interaction. Workshop on Simulation and Interaction in Virtual Environments. Iowa City, IW: ACM Press.