

VIRTUAL REALITY FOR TRAINING: EVALUATING RETENTION OF PROCEDURAL KNOWLEDGE

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

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

ABSTRACT

We conducted a study that contrasted the effects of virtual reality versus conventional computer-based instruction on retention of a simple procedure involving equipment operation. Seventy-two subjects between the ages of 20 and 30 were recruited from the general public. An intelligent tutor was used to manage the instruction in both treatments. The results were consistent with earlier research that reports that VR may not be superior to conventional electronic media for training certain intellectual skills. Implications for future research are discussed.

1. 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. Although research on skill decay has established practice as a strong predictor of skilled performance (Arthur, Bennett, Stanush, and McNelly, 1998). Little is known about how training that makes use of authentic contexts, such as those found on the job, is related to retention. There is some evidence that skill retention is related to the degree of resemblance between the learning context and the context in which performance is evaluated. The contextual attributes of virtual reality (VR) have been exploited in compelling ways in clinical psychology (Rizzo, Buckwalter, Neumann, Keselman and Theibaux., 1998)., task analysis (Martin., Sheldon, Kass, Mead, Jones and Breaux, in press). and practice environments such as military simulation (Knerr, Lampton, Witmer, Singer, Parsons and Parsons, in press; Lochlan, 1997). By engaging a person in what appears to be an authentic context, these studies suggest that affective or cognitive behaviors can be developed with VR at least as well as with conventional methods. There is little evidence, however, that behavioral change resulting from VR exposures are more durable than changes obtained without VR.

Despite VR's potential for simulating the real world, however, a framework for guiding the selection and design of VR-based training has not been developed and much work remains to be done before the training potential of VR is fully known. Little is known about how, or if, VR is superior to other forms of 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 (Clark, 1994) 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. For example, the acquisition and retention of

knowledge of a procedure involving simple actions such as pushing buttons or turning valves may not benefit more from a realistic simulation than from, say, a verbal description accompanied by photographs. We would not argue that simulations that afford authentic practice facilitate skills involving psychomotor learning. There is substantial evidence that authentic practice is a predictor of skilled performance (Ericsson). Few studies, however, have examined the degree to which specific types of knowledge are more or less acquired with VR than with, say, instruction delivered by a mouse and windows interface.

As a first step toward evaluating the instructional utility of VR, we decided to evaluate whether its attribute for simulating the spatial context of the real world was related to retention of procedural knowledge. We chose multimedia-based instruction as the baseline because it is commonly used for delivering instruction that makes use of simulations, is cheaper to develop than VR-based instruction, and contrasts sharply with VR in the portrayal of procedural sequences involving spatially-distributed events.

2. An experiment on knowledge retention

This paper reports the outcome from a study that addressed the effects of alternate presentation modes on knowledge retention. The primary purpose of this study was 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 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 full freedom of 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 recall more of the procedure over time than would subjects who learned the procedure from a more abstract presentation. The learning goal of the courseware was recall of a sixteen-step procedure for operating several devices distributed throughout several rooms in a naval ship. The devices included equipment consoles and pipe valves associated with a gas turbine engine (GTE) used by the U.S. Navy on some of its ships. Subjects were seventy-two people hired through a temporary employment agency. All subjects were at least high school graduates and almost all subjects were between the ages of 20 and 30. There were 42 subjects in the 2D group (26 females, 20 males), and 26 subjects in the VR group (10 females, 16 males).

The research was conducted on a Silicon Graphics computeran IRIX 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) with a 60 degree field-of-view, and three Flock of Birds trackers and an extended range transmitter 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 (Horwitz, Fleming, Regian, and Stiles 1996; Stiles, McCarthy and Pontecorvo, 1995). The 2D human-computer interface, on the other hand, consisted of a conventional windowing interface 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. VIVIDS is based on its predecessor, RIDES[®] (Munro, in press; Munro, Johnson, Surmon and Wogulis, 1993). 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. Selecting a machine on the floor plan with a computer mouse opened a separate window showing a close-up view of the device. Subjects used the computer mouse to interact with buttons on the machine console and other features of the device. The instruction prompted the subject to open windows and interact with devices 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 her head. Movement through the rooms 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 areas of a floor plan of the ship. In contrast, VR subjects were merely instructed to move to devices in the ship.

The instruction addressed two learning objectives. The first objective addressed recall of the location of the objects included in the procedure. There were sixteen objects that subjects had to locate, including three control consoles, five button panels, six buttons and two pipe valves from among multiple instances of each. The second objective was for subjects to locate the objects in a specific sequence. All of the objects were distributed throughout several rooms and passageways in the ship.

The instructional strategy provided by the software consisted of verbal prompts (text-to-speech) that directed the subject to locate devices in the procedure. Remediation was provided when the subject selected devices not included in the procedure or in the wrong sequence. 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. If a subject made two errors in a row, the software restated the instruction for that step of the procedure. The same information could also be obtained by clicking on a "Don't Know" button. Subjects in the 2D condition were instructed to use the "Don't Know" button if they were unable to determine what to do next. Subjects in the VR condition were instructed to state aloud if they didn't know what to do next, and the proctor played the "Don't Know" button for them.

The instruction was conducted in two phases. The first phase instructed subjects to touch each device in the procedure. To facilitate visual searches, devices were highlighted by alternating their colors in the manner of a flashing beacon. Flashing terminated when the subject selected a device. A few of the devices in the VR tutor were indicated by large arrows so that subjects could locate them from a distance. Following the first phase of instruction in which subjects were led through the procedure only once, the second phase prompted subjects to practice the procedure three times without the benefit of visual cues.

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 the "Don't Know" button, at which time, verbal instruction about the location of the device was supplied. Subjects repeated the procedure three times. Each time a subject completed the procedure, the computer stated the number of devices correctly selected (e.g. "...your score was 12 out of 16").

3. Procedures

Each subject followed the same general plan: (1) orientation to the task, (2) training with the interface, (3) interaction with the courseware, (4) a five-minute post-test immediately after, followed by (5) a five-minute post-test ten days later. Orientation to the task consisted of informing the subject about the nature of the procedure to be learned. Training procedures addressed use of the human-computer interface and were substantially longer for the VR group than for the 2D group. The VR group spent 40 minutes learning how to use the VR interface, but the 2D group only practiced with the mouse and windows interface for 4 minutes. Little practice was necessary for the 2D group because most of the subjects were familiar with windows-and-mouse-based interfaces. The rationale for enabling the VR group to practice using the interface was based on a pilot study we conducted earlier (Hall, Stiles and Horwitz, 1997), in which we observed that novices required at least 30 minutes to develop a facility for using the virtual reality gear. In this experiment, only two of the subjects reported prior experience with virtual reality headgear and gloves. Subjects first observed a demonstration showing how to use the VR gear, then they were taught to navigate around a virtual equipment room that was different from the rooms used in the experiment.

Instructional time was allowed to vary freely. Our pilot study indicated that the facility with which subjects manipulated the human-computer interface was a primary source for the variance in instructional time. Unlike subjects in the 2D condition for whom all devices could be quickly located on a floor plan or via parent objects, VR subjects viewed objects in the virtual world the same way they would have viewed objects in the real world - by walking among them. We had observed during the pilot study that even the most capable subjects occasionally got lost or could not locate a target object. Searches sometimes accounted for much of the time spent traversing the virtual world. To minimize time spent in the VR tutor for this study, proctors informed subjects when they were headed in the wrong direction or if they moved away from an area in which they had not completed a procedure. The proctor-supplied admonishments were, in most cases, identical to those supplied by the computer-based tutor, e.g. "That is incorrect." To remain consistent with the tutor's instructional strategy, the proctor pressed the "Don't Know" button when the subject made two consecutive errors. In some cases, however, subjects were unable to determine how to proceed from an incorrect path and the proctor had to supply more information, e.g. "you need to turn yourself around 180 degrees." Proctors only supplied the minimum information necessary to recover the instructional path, but did not lead subjects to objects in the procedure.

Proctors recorded observations of the level of facility that subjects exhibited with using the human-computer interface and any problems encountered to determine whether any systematic problems or patterns of interaction were suggested. We also recorded the time that subjects spent with the instruction, number of rest breaks taken and subjects' scores for each of the three practice drills during the instruction. Achievement was tested using a 1/10th scale model of the devices included in the instruction. Subjects

were tested for identification of the correct devices in the procedure, for the sequence of steps in the procedure, and had to recall the names of the devices. The test was administered to individual subjects in a private room. The model was elevated to enable viewing it at eye-level. Subjects were first walked around the model before starting the test. They were then instructed to point to each of the devices in the procedure, in the correct sequence and to state the names of the objects. Two proctors recorded each subject's progress while observing the subject execute the procedure. A printed floor plan of the model was used to record observations. There was only one correct path for executing the procedure. The same model was used for the immediate posttest and the delayed retention test.

4. Results for Skill Decay

A quantitative contrast between the 2D and VR groups was performed for recall of the devices in the procedure (Table 1) and for the sequence of steps in the procedure (Table 2). Over a ten-day period between tests, average knowledge retention between the two groups on both types of knowledge, was about the same. A repeated-measures ANOVA did not find a statistically significant difference in the contrast of scores between 2D and VR ($p < .05$). Subjects were asked to state the name of each device while they were tested on the procedure. There were 29 words associated with the devices in the procedure. On average, subjects could recall only 3 or 4 of the words during the tests. The lowest score was zero and the highest score was 12.

	Immediate		Delayed		Diff
	M	SD	M	SD	
VR	85	13	69	18	16
2D	71	24	60	19	11

Note: Tests were graded on a 100-point scale.

Table 1: Test Results for Object Identification

	Immediate		Delayed		Diff
	M	SD	M	SD	
VR	88	15	76	18	12
2D	69	26	54	19	15

Note: Tests were graded on a 100-point scale.

Table 2: Test Results for Step Sequence

5. Other Observations

On average, VR subjects spent more than twice as much time in the instruction ($M = 69$ min., $SD = 20$ min.) than did the 2D group ($M = 29$ min., $SD = 8.7$ min.) However, only two VR subjects reported brief and very mild nausea. All subjects experiencing discomfort or fatigue recovered after short breaks during the instruction. While none of the 2D subjects took breaks, four VR subjects took no breaks, eleven took one break and eleven took two breaks. Breaks ranged from two to six minutes.

Twenty-five of the twenty-six VR subjects received assistance with orientation or getting lost. About half of the VR subjects had some difficulty effecting actions with the PINCH gloves. Subjects with small hands had the most difficulty connecting the gloves' fingertips to effect actions while subjects with large hands tended to expand the gloves so that the electrical contacts would separate.

When rest breaks were factored out, variance in time spent with the instruction was related to subjects' accuracy in performing the procedure, versus guessing. There was a statistically significant negative relationship between practice drill scores and time spent in the instruction, $F(1,70) = 34.75$, $p < .001$. Practice drill scores

explained 32% of the variance in time. The lower the practice drill scores (by selecting incorrect objects), the longer the time subjects spent in the instruction. Both groups scored nearly the same for each of the three practice drills, suggesting a similar rate of increase in skill during instruction. Practice drill scores are presented in Table 3.

	Drill 1		Drill 2		Drill 3	
	M	SD	M	SD	M	SD
VR	53	13	73	14	84	14
2D	58	16	69	16	79	15

Note. Scores are based on a 100-point scale.

Table 3: Practice Drill Scores

6. Discussion

This study attempted to contrast two types of media on knowledge retention. There was no statistically significant difference in skill decay. While the absolute performance of the VR group was generally higher on tests of object identification and step sequence, most of that difference can probably be explained by time spent in the instruction. On average, the VR group spent more than twice as much time engaged in the instruction as did the 2D group. VR subjects spent more time moving between objects in the procedure than did 2D subjects, and spent more time searching for correct objects after selecting incorrect ones. Another result supporting the lack of a difference between tutors was in the similarity of practice scores between groups. Both groups appeared to learn at similar rates across practice trials, suggesting that, for recall of the procedure used in this study, the educational utility of both tutors was about the same.

The generally low recall of device names by both groups was not surprising because subjects were not instructed to remember the names of objects. Although the instruction did refer to objects by name when prompting subjects to proceed to the next station in the procedure, subjects were generally unable to recall device names while they identified objects on the model used for the posttest. Perhaps if we had used a multiple choice test to prompt recall of device names, subject might have recalled more.

We were surprised by the stamina exhibited by the VR subjects, all of whom engaged in the instruction while standing up. The average time spent immersed was almost 70 minutes, which contrasted sharply with the 45 minute maximum during the pilot study in which subjects were seated, were immersed in a much simpler world, but generally expressed fatigue. We had anticipated that standing subjects would experience at least some motion sickness and fatigue. It is possible that the remuneration provided to subjects for participating in the study motivated them to overcome physical challenges. All 72 subjects were paid nearly \$100 each, with most of the compensation awarded after completion of the retention test. In contrast, subjects during the pilot study were unpaid volunteers.

Technical issues with the VR tutor may have had an impact on the study. Despite our attempts to minimize differences between the tutors, the VR environment imposed some significant challenge on subjects. These challenges included the visual distance between subjects and objects in the virtual world, and navigation.

Visual distance:

2D subjects viewed all devices from a fixed distance – the distance portrayed in pictures of the devices. VR subjects, on the other hand, viewed objects in the round from any distance. However, VR subjects had to interact with objects in a way that did not afford a view of the broader context. VR subjects had to get within inches of an object in order to touch it with the PINCH glove. By forcing VR subjects to work in close quarters, the views afforded of the devices in the VR tutor may have been rendered more impoverished than the views afforded in the 2D tutor. The sixty-degree view afforded by the HMD may also imposed a kind of tunnel vision, effectively blocking out peripheral views.

Navigation.

VR subjects had to spend time walking to stations in the procedure which took much more time than moving a mouse between stations in the 2D tutor. While it was unclear if VR subjects expended more mental effort learning the procedure, they did have to spend more time attempting to locate target objects from among other objects that populated the rooms and hallways in the environment. Most VR subjects were also distracted by difficulties with effecting actions with the PINCH gloves. They sometimes had to pause while proctors adjusted the gloves. The extended time spent dealing with navigation, search and with the interface may have precluded VR subjects from acquiring a more robust memory of the procedure.

Way finding was also a problem for VR subjects. Almost all VR subjects received some navigational assistance from proctors during the experiment. The problem may have been rooted in the fact that target objects were difficult to see from a distance. Although the tutor provided navigational information during the initial pass through the procedure and during remediation, subjects often were unable to successfully follow the navigational prompts. Some objects were located behind other objects or around a corner in another hallway. Problems arose when a subject became lost while searching for the next object. The tutor provided route instructions before the subject set off for the next station, but subjects were almost always observed to be looking in the wrong direction while instructions were being played. By the time a subject was looking in the correct direction, the instructions would cease, and then the subject had to recall the instructions from memory while relating them to the new context. The result was usually for the subject to ask for the instruction again, or get lost while searching for target objects. Subjects were also observed to forget the navigational instructions while moving to the next station in the procedure. Playing the "Don't Know" button, however, resulted in a repetition of navigational directions from the visual perspective of the last station. The VR tutor was not designed to incorporate the subject's current perspective into directional prompts. For the most challenging paths between objects, instructions were automatically repeated several times so that subjects would hear the instructions after they had set off in the direction of the next station. We also provided artificial landmarks such as large colored arrows that pointed to some of the objects in the procedure. Despite these strategies, subjects generally failed to utilize them with much success during their first pass through the procedure. Only during the three subsequent passes through the procedure did some subjects begin to navigate skillfully between stations.

Authoring the instruction that would provide directional advice from any point in the world would be impractical because there are an infinite number of visual perspectives in a procedural path from which a subject could request route instructions. There is the possibility of using a subject's position and viewpoint to mathematically derive navigational prompts, but much information would still have to be authored for instructing learners how to position themselves to skillfully operate devices. During the study, many subjects were observed to position themselves relative to the device in ways that made for awkward manipulation of that device. To address some of these problems, we are investigating the use of an intelligent pedagogical coach that has human form and can lead the learner around the environment while demonstrating how to operate devices in the virtual world. By having the agent available as a guide, much of the overhead associated with verbalized description can be reduced.

We are not convinced that, for learning simple procedural knowledge, an authentic context would be superior to more abstract presentations. Past media studies, independent of instructional strategies, have not shown significant differences between types of media for learning skills. For example, Regian (1997) 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 (1996) 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.

7. 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 conventional windows and mouse interfaces will impact learning 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

- Arthur, W., Bennett, W., Stanush, P. L. & McNelly, T. L. (in press) Factor that influence skill decay and retention: A quantitative review and analysis. *Human Performance*.
- Clark, R.E. (1994). Media will never influence learning. *Educational Technology Research and Development*, 2(42), 21-29.
- Ericsson, K. A. & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, (47), 273-305.
- Horwitz, C., Fleming, J., Regian, J., & Stiles, R. (1996). Software tools for embedding principled instruction in virtual environments. In proceedings of the *IMAGE Conference*, Scottsdale, Arizona 23-28 June 1996.

Knerr, B. W., Lampton, D. R., Witmer, B. G., Singer, M. J., Parsons, K. A. & Parsons, J. (in press). Recommendations for using virtual environments for dismounted soldier training. . In proceedings of the *Interservice, Training, Simulation and Education Conference*, Orlando, Florida, 20 November to 3 December, 1998.

Lochlan, M. E. (1997) Virtual reality simulator (VRS) for training ship handling skills. In *Virtual Reality, Training's Future?* (pp. 19-29). Plenum Press: New York.

Martin, K. M., Sheldon, E., Kass, S., Mead, A. M., Jones, S. A. & Breaux, R. (in press). Using a virtual environment to elicit ship-handling knowledge. In proceedings of the *Interservice, Training, Simulation and Education Conference*, Orlando, Florida, 20 November to 3 December, 1998.

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

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

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.

Rizzo, A. A., Buckwalter, J. G., Neumann, U., Keselman, C., & Theibaux, M. (1998). Basic issues in the application of virtual reality for the assessment and rehabilitation of cognitive impairments and functional disabilities. *CyberPsychology & Behavior*, 1(1), 59-78.

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

AUTHOR CONTACT INFORMATION

Craig R. Hall
Air Force Research Laboratory
AFRL/HEJT
7909 Lindbergh Dr.
Brooks AFB, TX 78235-5352
Office: (210) 536-1989, Fax: (210) 536-2902
Email link: <mailto:hallc@alhrt.brooks.af.mil>

Carol D. Horwitz
Command Technologies, Inc.
AFRL/HEJT
7909 Lindbergh Dr.
Brooks AFB, TX 78235-5352
Office: (210) 536-2764, Fax: (210) 536-2426
Email link: <mailto:cdh@count.brooks.af.mil>