Training in Virtual Environments: Experimental Evaluations and Implementation Strategies

Barbara Buck and Bruce Perrin
The Boeing Company
St. Louis, MO
barbara.j.buck@boeing.com, bruce.m.perrin@boeing.com

ABSTRACT

This paper summarizes a series of research studies conducted over a period of six years in which we investigated the effectiveness of virtual environments for aircraft maintenance training. In these studies, we systematically evaluated numerous aspects of virtual maintenance trainers (VMTs) that may impact training effectiveness, including comparisons of immersive vs. desktop virtual environments; high vs. low detail graphics; task selection and specific tasking requirements; and the effects of individual differences on learning effectiveness within virtual environments. In each study, we conducted an evaluation of training effectiveness, collecting objective performance measures of declarative knowledge, training transfer task performance time, and transfer performance errors. In addition, individual difference measures of spatial reasoning aptitude, computer/video game experience, and hand tool experience afforded the opportunity to analyze the impacts of these variables on VMT effectiveness. Within each study, rigorous experimental protocol in training procedures and data collection remained constant, allowing us to compare results across the multiple studies. The paper highlights significant results from each of these individual studies, as well as generalized findings across all of the studies. Our studies showed that interactivity and high graphic detail are important for training effectiveness, and that desktop VE trainers were a significantly more effective medium than immersive training. In addition, individual spatial reasoning aptitude is a significant mediator of virtual training effectiveness when the task would be learned and performed with limited visual feedback in the real world (e.g., a tactile or "blind" task). The impact of the spatial processing aptitude is similar for both hardware and VE-based training, however, when the task is normally learned and performed in the visual field. This paper summarizes these findings, as well as other study results, as a series of empirically driven guidelines for the implementation of virtual environments for aircraft maintenance training.

ABOUT THE AUTHORS

Dr. Barbara Buck has had fifteen years of human factors experience at Boeing, both in applied research and advanced design applications. Her primary focus has been in the areas of training evaluation, workload prediction and assessment, modeling pilot performance and decision-making, cognitive ability assessment, performance measurement, and development of principles for decision aiding and display design. Her efforts have involved contractual and internal research and development, as well as direct support to aircraft programs. Currently, Dr. Buck is an Instructional Designer with Boeing's Maintenance Information Systems group, and is principal investigator on an internally funded research and development project evaluating the effectiveness of virtual environments for training.

Dr. Bruce Perrin is a Technical Fellow, assigned to the Maintenance Information Systems area of Boeing Aircraft and Missile Systems. Bruce has been employed at Boeing for 20 years, during which time he has been responsible for the analysis, design, development, and evaluation of training and decision support systems. He currently leads the company's research and development of Advanced Distributed Learning (ADL) technology and methods, and is principal investigator on the Advanced Distributed Learning Shareable Content Object Reference Model (ADL/SCORM) Independent Research and Development project. His research interests and publications include work in the areas of cognitive analyses of training environments, expert systems for training and job-aiding, training system evaluation, adaptive training methods, and decision-making and -aiding under uncertainty.

Training in Virtual Environments: Experimental Evaluations and Implementation Strategies

Barbara Barnett and Bruce Perrin
The Boeing Company
St. Louis, MO
barbara.j.buck@boeing.com, bruce.m.perrin@boeing.com

INTRODUCTION

Virtual environments (VEs) for aircraft maintenance training offer a broad range of potential advantages over their hardware-based counterparts. Significant cost savings may be realized during development by reusing existing engineering models. Virtual maintenance trainers (VMTs) are also more portable, require less space, and involve less costly hardware. Consequently, they can be fielded in greater numbers to enable higher levels of student throughput and to provide training whenever and wherever it is needed. Finally, system upgrades and modifications are more cost efficient, as VMTs are updated with software changes rather than by rebuilding a physical device. There are, however, also a number of potential disadvantages to VMT technology. These include a loss of fidelity in the training device and giving up the "hands-on" feel for the task being trained. It is important to consider the tradeoffs between these potential limitations and the possible advantages when implementing VMTs for aircraft maintenance training.

Since 1999, researchers at the Boeing Company have been evaluating these potential tradeoffs in order to provide empirically based guidance in the use of VEs for maintenance training (Perrin & Barnett, 2001, 2002; Barnett & Perrin, 2002). Training effectiveness in this series of studies was evaluated by comparing performance following VE-based training to that realized after training on a hardware mockup.

To establish an accurate tradeoff between VMT technology and hardware-based training, the relative training effectiveness of alternative implementations needs to be established. Within the realm of VMTs, there is a wide range of capabilities and features that might be used for training. For example, designers and developers of VMTs may choose to provide visual cues to the trainees what are not available in real task performance (e.g., the ability to see through solid walls). As another example, designers may select from varying levels of detail and resolution within the VE, as well as methods of presentation (immersive vs. desktop). The question is which of these VMT presentation methods will lead to the most effective training and how does that training compare to a hardware-based approach.

Finally, we have addressed the question of which tasks are best suited for VMT-based training. There are a broad spectrum of maintenance training tasks, ranging from simple remove and install (R&I), to troubleshooting and fault diagnosis--only some of which might be suitable for training with VMTs. Some of our research has addressed this question by changing the nature of the basic maintenance task investigated; others by conducting evaluation studies on an altogether different maintenance task.

The result of this multi-year research effort is a set of empirically driven guidelines for the implementation of VEs for maintenance training. This paper summarizes this program of research and the relevant findings as it pertains to VMT development and design.

EXPERIMENTAL BACKGROUND

In order to facilitate comparison of data across multiple studies, rigorous experimental protocols and procedures were put in place prior to the initial study and maintained throughout the entire series of studies. Aircraft design models were used to implement the VE and to build the hardware mockup, assuring consistency between the training media. In addition, identical written and audio instructions were presented to students across training conditions. Progression through the training was self-paced, and only a single pass through the materials was allowed. We collected data on identical measures of performance across these studies, including training time, scores on a knowledge test, and speed and accuracy in performing the maintenance action on the hardware mockup. Each of these procedures, as well as the maintenance task used for training, the test measures, and individual aptitude measures are described in detail below.

Basic Maintenance Task

The primary maintenance-training task used in this series of studies was the removal and replacement of an aircraft fuel valve. The 24-step procedure was selected as the test case because it provided an appropriate amount of complexity; it imposed both cognitive and physical demands on the trainees, and allowed for limited visual access within the actual aircraft and the physical mockup. In this task, an access door and an additional part must be

removed to provide physical access to the fuel valve (see Figure 1). Even after they are removed, the size of the access opening and the position of the valve limit visual access to fasteners that are removed during the task. This means that a large portion of the mockup task is done only by "feel". The visual constraints of this task provided an opportunity to showcase the unique ability of the VE to see though walls and to get "inside" the compartment for visual access that is otherwise unattainable in the hardware mockup or aircraft.

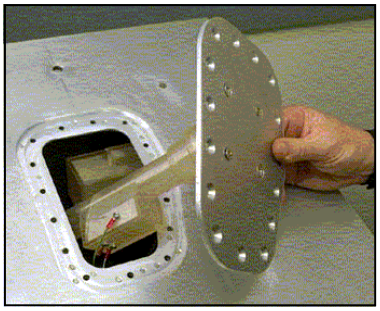


Figure 1. Fuel Valve mounted inside aircraft wing compartment.

Hardware Mockup

A full-scale hardware mockup was the primary device used for the baseline training condition, and was also used as the test device for training transfer assessments (see Figure 2). The aircraft wing section was built from aircraft production models and drawings, but did not contain actual parts. Instead, it used off-the-shelf hardware and simulated aircraft parts built to specification from epoxy resin. The performance of individuals trained using different implementations of the VE was compared to the performance of those trained on the hardware mockup.

Figure 2. Participant completing training in the hardware mockup condition.

Procedure

Participants were Boeing employees, summer college interns, and U.S. Navy aircraft maintenance trainees who volunteered to participate in the training studies. These volunteers were not pre-screened for skill or ability, although people who had previous experience with either aircraft maintenance or fuel systems (other than as a Navy trainee) were not allowed to participate in the study. Each participant was randomly assigned to only a single training condition. For purposes of this paper, we have pooled data from 250 participants over the course of 7 different studies.

Although every effort was made to assure comparability of our findings across studies by strictly adhering to a common procedure, there are threats to the validity of any analysis that draws on data collected over such an extended period of time. Primary among the threats in our case are potential differences among study samples based on history or due to the differential selection of participants (Cook, Campbell & Peracchio, 1990). The purpose of many of these studies, however, has been to provide replication of findings or to test specific predictions from previous research. Given the consistency with which our findings have been replicated and our predictions confirmed across studies, the impact of these factors is, most likely, slight. Nonetheless, the reader should take this issue into account in any interpretation or application of our findings.

At the beginning of the test session, participants completed a short background questionnaire in which they were queried as to their level of previous experience with hand tools and basic maintenance, as well as previous experience with 3D video games. This was followed by a paper and pencil test to assess their spatial visualization aptitude (Ekstrom, French, Harman, & Dermen, 1995). After the background information was collected, the participant was given instructions regarding the training procedure, then proceeded directly through the training. The training was self-paced, and the amount of time each participant spent in training was recorded.

Once the training was completed, participants completed a written test of their task knowledge. This assessment was comprised of questions about the maintenance procedure, and was used to assess how much declarative knowledge each participant acquired about the task during training. Some of the questions were detailed and specific, so that subtle differences in acquired knowledge could be identified.

The participants then completed the training transfer task. During this phase, participants from both groups performed the fuel valve remove and replace task on the full-scale hardware mockup without the aid of written or audio instructions for reference. They were asked to work as quickly and as accurately as possible, and to complete the steps in the order in which they were trained.

Training Effectiveness Measures

A number of training effectiveness measures were collected during the study, including knowledge test scores and transfer task completion times. We also measured significant errors that occurred in performing the task. These included omitted steps, incorrect actions, and forgotten procedures. In addition, time spent on the self-paced training was recorded.

In an effort to increase reliability of our measures and to minimize the potential impacts of speed/accuracy tradeoffs, we combined these three aspects of performance – training transfer error, task performance time, and knowledge test score – into an aggregate of combined performance. This was achieved by converting each individual measure into a z-score for the sample population, and summing across the three measures for each individual to derive a single index of combined performance. This measure will be reported as the primary index of performance.

Training Manipulations

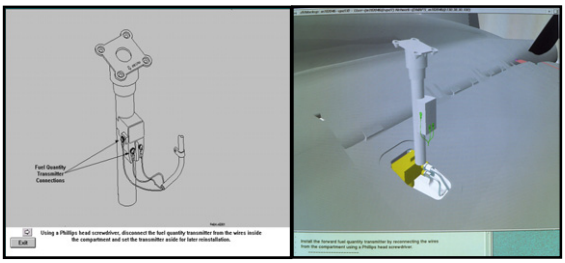
Across the series of studies, we had three primary goals. The first goal was to validate VE training effectiveness relative to the baseline hardware mockup training. The second was to optimize VE cues for learning (i.e. present information within the VE to maximize learning). The third goal was to determine when VE training is most effective (i.e. determine if certain tasks are better suited to learning by VE than others). In order to provide empirical guidelines relative to these goals, we implemented two primary types of experimental manipulations. These were: manipulations to the virtual environment and manipulations to the maintenance task characteristics. Results in each of these areas are described in the following sections.

TRAINING STUDIES AND RESULTS

Level of Detail and Interactivity

In one of our initial studies, we investigated the impact of levels of training interactivity and level of graphical detail on training effectiveness using Desktop VE. Two levels of VE detail were combined with two levels of training interactivity to make four separate VE training conditions.

VE Training was presented in two types of training environments. The *Low Detail VE* condition used line drawings, illustrated in Figure 3(a), which were taken directly from the existing Interactive Electronic Technical Manual for this task. While these drawings would not generally be classified as a VE, they do provide a 3-D perspective, but only supported by the most rudimentary Gestalt principles such as Closure and Good Continuation (Garner, 1974). The relevant parts were highlighted with color, and animation was used to illustrate the actions within a step. The *High Detail VE* condition consisted of three-dimensional models that were embellished with color, highlighting, shading, texture, and animation. These drawings, illustrated in Figure 3(b), were more realistic than the line drawings presented in the Low-Detail condition. Also, the High Detail condition allowed for different camera views, enabling the trainees to "get into the compartment" for a better view of the aircraft parts in question. The VE images were animated to show movements of parts throughout the training task.



(a) (b)

Figure 3. Sample VE training under the (a) Low Detail and (b) High Detail conditions.

Two types of interactivity were employed in this study. One level was *passive*; that is, the trainee did not interact directly with the training material except to advance to the next step in the procedure. Trainees simply watched and listened to the training material. *Interactive* training required the trainees to select the relevant parts to be removed, replaced, etc., to initiate animation, thus making the trainees active participants in the learning process.

Results

The possible impact of different VE formats on training effectiveness was evaluated with analysis of covariance, using previous tool experience as the covariate. Significant differences were obtained on the combined performance scores between the two levels of detail ($F = 10.01$, $p = 0.0033$). Illustrated in Figure 4, trainees in the high-detail VE condition scored higher in combined performance than the low-detail VE trainees. The trend for better performance in the high-detail VE condition was also found in the analysis of the individual performance measures. The trainees in the high-detail VE condition made significantly fewer errors than trainees in the low-detail VE condition ($F = 5.41$, $p = 0.0261$).

A significant effect was also found in the interactivity manipulation. The Interactive training condition produced significantly faster performance times than the Passive training condition ($F = 6.44$, $p = 0.0159$). Finally, a significant interaction was found between level of detail and interactivity manipulations. The Interactive VE

trainees scored significantly higher on the Knowledge Test than trainees in the other three computer-based training conditions ($F = 4.77$, $p = 0.0359$).

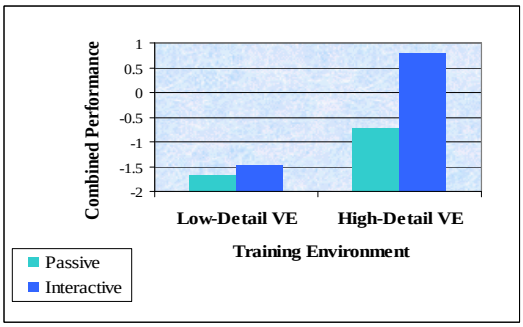


Figure 4. Combined performance scores.

Collectively, these results suggest that interactivity and higher levels of detail are important for effective VE-based training. For most measures, performance in the Interactive VE condition did not differ from that of the mock-up training condition. Illustrated in Figure 5, combined performance for the Interactive VE condition approaches that of the mock-up training condition.

With a finding of "no statistically significant difference", one might be tempted to conclude that VMTs are fully as effective as hardware-based training, and discontinue these studies. However, we believed that there were several reasons to proceed. First, as evidenced by Figure 5, there is still room for improved training effectiveness within the VE. Second, the direction of this difference, while not statistically significant, has been replicated several times in later studies. The consistency of this result suggests that the lack of statistical effect may be more a matter of sample size than a lack of a true difference. And third, inspection of the relative variances in performance of VE and hardware-trained individuals indicates that VE-based training has a significantly more inconsistent effect on learning.

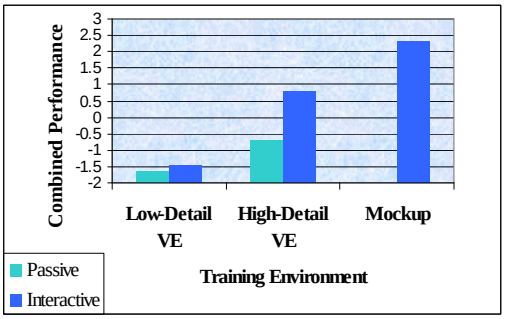


Figure 5. Combined performance – CBT vs. mockup.

Quite simply, some individuals learned effectively using the VMT; others did not. Consequently, part of the lack of statistical difference may be due to differences in variance between the groups. We will return to this difference in variance and the possible source of it in the section on selecting tasks for VE-based training. Given these factors, however, we considered it important to continue our investigation into methods to improve VMTs.

In this quest, one of the follow-on issues we examined was the difference in the amount of time that participants were exposed to the training. Average training time with the mock-up was approximately three times as long as average training time on any of the VE training conditions. It was not clear whether the amount of exposure or the type of exposure caused the differences in performance. To test this hypothesis, we added another training condition where training time for individuals in the high-detail VE training condition was matched to training time for individuals in the mockup condition.

Interactive VE, Matched Training Time

In the Interactive VE Matched for Training Time condition, participants completed the Interactive VE training course repeatedly for an average of 35 minutes. Most participants completed the course three or four times within that time period. Ten participants completed this training condition.

Results

Combined performance scores were analyzed for the new test condition, and were compared to VE Interactive and mock-up scores from the previous portion of the study (Figure 6). The results show that, for combined performance scores, additional training time with the Interactive VE had no effect on transfer task performance. This trend was the similar for the analyses of each individual performance measure as well.

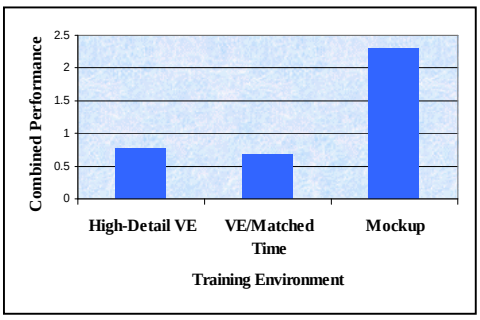


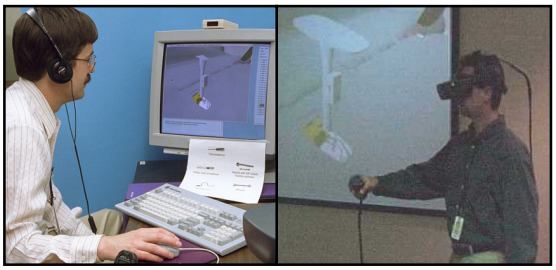
Figure 6. Combined Scores – Matched training time had no effect on performance.

Based on these findings, we concluded that additional exposure to the interactive, high-detail VE training environment did not impact training effectiveness. Even working through the training an additional 2-3 times did not improve training transfer performance. These findings

suggest that factors other than training time alone resulted in the performance differences observed between VE and mockup training.

Mode of VE Training: Desktop vs. Immersive

In this study, Immersive VE training was evaluated against the High-Detail Interactive Desktop VE and Hardware Mockup training. The Desktop VE (Figure 7a) condition consisted of the same high-detail, interactive environment described in the previous study. While desktop VE systems are limited in the depth cues that are available from 2-D displays (such as shading, texture gradients, and monocular motion parallax), the Immersive VE condition presents additional depth cues. Specifically, such systems provide stereoscopic information, presented using a head-tracked visor, as illustrated in Figure 7b. Trainees are able to control field of view by adjusting their head position within the environment, and are able to select parts and tools within the environment using a 3D mouse. When the 3D mouse was in the participant's field of view, a "virtual hand" appeared within the VE that could be used for grasping tools or manipulating parts. Training instructions were presented through auditory instructions, and could be repeated as necessary.



(a) (b)

Figure 7. Two training formats: (a) Participant training on the VE Desktop trainer; (b) Participant using the immersive visor and 3-D mouse.

Prior to the study, an extensive series of pilot tests was conducted to optimize the 3-D mouse interface to the Immersive VE for training purposes, thus minimizing the complexity of the trainee's interaction with objects. These VE interface implementations allowed the trainee to become proficient rapidly within the immersive environment, eliminating the need for lengthy familiarization training.

Results

The results for the Immersive vs. Desktop VE vs. Mockup study are summarized in Figure 8. As illustrated, combined performance for all 3 training conditions differed significantly from one another. Combined performance was highest for the mockup condition, followed by the Desktop VE. Performance was poorest for the Immersive VE condition ($\boldsymbol{F}$ (2,59) = 13.12, p<.0001). A similar pattern of results was revealed for each of the individual performance measures of test time ($\boldsymbol{F}$ (2,59) = 4.99, p=.0102), errors ($\boldsymbol{F}$ (2,59) = 6.05, p=.0042), and knowledge test scores ($\boldsymbol{F}$ (2,59) = 25.40, p<.0001). For each measure, performance followed the same pattern, where mockup performance was best and Immersive VE was the least effective.

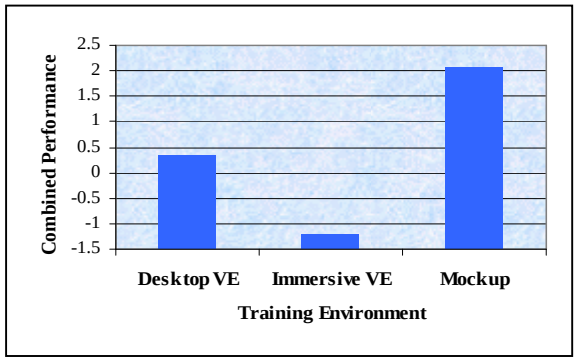


Figure 8. Training effectiveness was best for the mockup condition; worst for the Immersive VE condition.

Immersive VE with Extended Practice

In an effort to explain the relatively poor performance of individuals trained with the Immersive VE, we speculated that familiarity with the unique aspects of an Immersive VE might be a potentially confounding issue. A number of trainees had difficulty working within the interface, despite the fact that every effort was made to optimize it. Some had to be repeatedly coached through the task. Given this limiting factor, we added an additional Immersive VE training condition to a follow-on study. In this condition, a separate immersive VE familiarization session was added to the training. In this extended practice condition, participants completed the basic familiarization, as well as an extra practice session within the immersive environment on a distinct maintenance task. This extended practice task used the same VE navigation and manipulation conventions, giving participants more practice within the immersive environment. The maintenance task used in extended practice had no similarities to the training task used in the study. Participants in the basic practice condition received an average of 10 minutes familiarization prior to the start of their training, while those in the extended practice condition received an average of 30 minutes of familiarization and extended practice. For the training effectiveness portion of the study, conditions in both training groups were identical.

Results

Training effectiveness data from the extended practice condition were compared with those from the original immersive condition. Illustrated in Figure 9 below, the additional practice did not alleviate the disparity between Immersive VE training and the other training conditions. Combined performance for both Immersive VE conditions

was still significantly worse than the Desktop VE and the Mockup training conditions (***F*** (2,59) = 10.66, $p<.0001$).

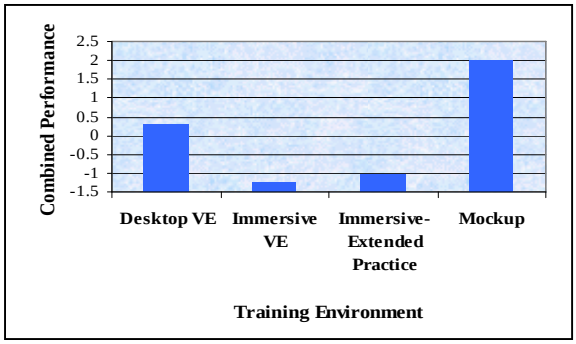


Figure 9. Combined performance not improved for the immersive extended practice condition.

Based upon the results of these two studies, we have concluded that for this specific maintenance task, Immersive VE is not an effective training technique. We then turned our efforts to determine if there were additional enhancements that could be made to the Desktop VE training conditions that would improve VE training effectiveness relative to the Mockup training condition.

Enhanced Interactive Desktop VE

Results of previous studies suggested that High-Detail Interactive Desktop VEs were most comparable to hardware mockup training in terms of training effectiveness. Yet despite the high training transfer performance relative to the low-detail, non-interactive and Immersive VE conditions, performance for the high-detail, interactive desktop VE was still not as good as that of the traditional hardware mockup training. For this study, we made a number of enhancements to the original high-detail desktop VE condition aimed at optimizing VE effectiveness. A number of these changes were based upon comments made by trainees in previous studies as to what was and was not effective in the current training. Other changes were based on an extensive analysis of the errors made by VE training participants on the training transfer task and on the knowledge test.

Improvements that were made to the original software included the addition of a virtual hand during certain scenes to provide size perspective for access doors and aircraft parts. We also improved text and audio instructions to indicate important steps and more explicitly reflect necessary actions. Resolution within the VE was improved, and colors were added to more accurately represent mockup parts. In addition, the aircraft skin covering the compartment and the access door were made slightly transparent to enable the student to see the underlying parts initially. Examples of current VE training screenshots are provided in Figure 10.

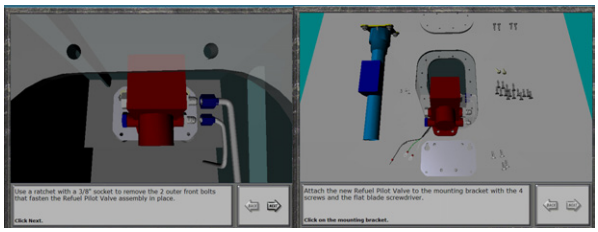


Figure 10. VE software screen shots.

Results

The comparison of combined performance was not significantly different for the two training conditions ($F(1, 39) = .18$, *n.s.*). This measure summarizes the impact of our VE training enhancements on overall training effectiveness, indicating that training transfer performance between the Enhanced VE and mockup was similar. Of course, this lack of statistically significant difference has been documented before. What proved new here were the differences among individual performance measures.

Turning to these individual measures of training transfer effectiveness, we observed no significant impact of training condition for our measure of test time ($F(1, 39) = 1.88$, *n.s.*). Time to complete the training transfer task was not related to the condition under which the participant trained. For errors, there was a significant main effect of training condition. While error rates for both training conditions were very low (less than 1.0), VE-based trainees actually made significantly fewer errors than those who trained with the mockup ($F(1, 39) = 14.25$, $p = .0006$). This difference is illustrated in Figure 11 below.

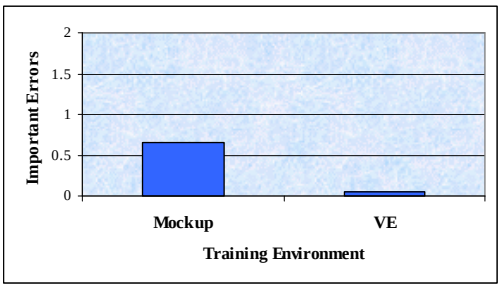


Figure 11. VE trainees made fewer training transfer errors than mockup trainees.

We also found a significant main effect of the knowledge test measure ($F(1, 39) = 9.40$, $p = .004$). For this measure, those participants who trained in the mockup condition scored significantly higher on the test of procedural knowledge than those who trained in the VE. This finding is illustrated in Figure 12.

Although VE trainees performed worse on the knowledge test, they did perform better on the transfer test, as measured by the number of errors they committed. This contradicts previous studies, where mockup performance was better than VE-based training for most metrics.

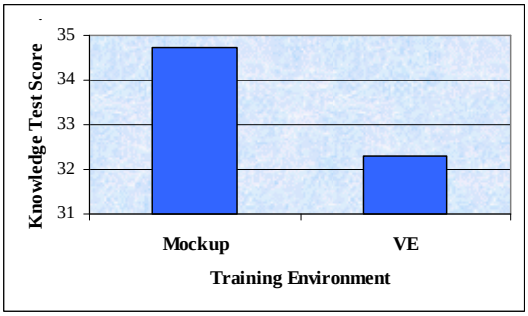


Figure 12. Mockup trainees scored higher on the knowledge test than VE trainees.

Results of this final study indicate that VE training holds significant promise in providing effective training and task rehearsal for a basic maintenance task. These findings suggest that the enhancements made to the VE for the latest study were effective in improving training transfer performance. In most cases, training transfer performance for the enhanced VE training was better than original VE conditions, was highly comparable to that of the mockup training, and, for one measure, performance following VE-based training was even better than mockup training. It was only the knowledge test that indicated better performance for the mockup training over that of the enhanced VE. Given that this test focused on very specific details of the fuel valve R&I task (e.g. size of fasteners, orientation of parts) in addition to procedural questions, it is not surprising that certain details were not observed within the VE training conditions.

Looking at the data for errors and for test time, it appears that any lack of declarative knowledge regarding the maintenance-training task failed to transfer task performance. This distinction is important, realizing that the focus of maintenance training is performing the task, not necessarily answering detailed questions about it.

Manipulations of Task Selection

Focusing now on the third goal of this series of studies, we sought to investigate which tasks are more suited for VE-based training, and to determine if there are empirically grounded guidelines for implementing VE-based training for selected tasks and not for others.

One significant finding in the data from our original study was significant differences in variability among training conditions. Specifically, variance measures for each of the VE-based training conditions was significantly higher than for that of the mockup condition. Descriptive statistics from the study indicated that variability in post-training performance for the VE-based participants was quite high compared to the hardware-trained individuals. Variance following training based on VE-Low Detail was the highest, with variance resulting from mockup-based training the lowest (see Table 1).

Table 1. F-test on performance variance following different types of maintenance training.

Training Environment	Variance	F-test
VE-Low Detail	7.79	**F** (3,19) =
VE- High Detail	6.83	12.98,
Mockup	0.60	$p < .001$

While these findings suggested that desktop-VE maintenance training would, in practice, have a substantially less consistent effect on learning than hardware-based training, one concern was that the participants in this study might not accurately represent the maintenance workforce. Consequently, we replicated this study with active duty Navy maintenance trainees at the Naval Air Station – Pensacola, FL. The results were unchanged. The post-training performance variance for Navy students trained with the hardware mockup was 2.01. It was more than 14 times greater for students trained on the low-detail VE at 29.16, while it was over 16 times greater for students trained with high-detail VE at 32.67. Combining data from the Boeing and Navy samples to yield a stable estimate of these parameters, the variances in learning scores following low and high-detail, desktop VE training both significantly exceeded the same variance following training on the hardware mockup ($F = 10.15$, $p < .01$ and $F = 8.95$, $p < .01$ for the low- and high-detail VEs compared to hardware training, respectively).

We also evaluated the possibility that VE-based training might have a more consistent within the immersive, high-detail VE. Results, however, were, once again, unchanged. The variance in post-training performance on the knowledge test was 31.84 compared to 2.62, a difference that is statistically significant ($F = 12.15$, $p < .01$).

Individual Difference Predictors

Parallel to finding these differences in variance between hardware and various forms of VE-based training, we sought to identify which, if any, of the individual difference predictors correlated with performance. The identification of these factors may help in understanding why VE-based training resulted in more highly variable performance compared to hardware-based training. We examined the three primary individual factors: prior experience with tools, past exposure to 3-D computer environments, and the spatial visualization aptitude. Our measure of prior experience with tools was generally a predictor of performance time, in that those with more prior experience were faster in performing the transfer task. This effect was consistent across training conditions. Past exposure to 3-D video games has frequently been cited as possible determinant of an individual's predisposition to learn from

a VE. Our research, however, has failed to find any consistent pattern between experience with various types of 3-D environments and post-training performance (Barnett & Perrin, 2002).

One factor in our studies, however, has consistently been found to correlate with performance following VE-based training. That factor is the individual's spatial visualization aptitude. This measure maintained a consistent relationship to learning and transfer performance in multiple studies across a number of our VE training conditions. Table 2 summarizes the correlations between visualization and knowledge test scores for one of our previous studies. The Low Detail condition produced post-training performance that was significantly associated with the visualization aptitude (r = .52). With increasing visual detail, the correlation between post-training performance and this aptitude was no longer significant (r = .22), and it dropped to near zero (r = .02) when training was provided on the physical mockup.

Table 2. Correlations between knowledge test scores and the visualization aptitude.

Training Environment	Correlation
VE-Low Detail	0.5203, p<.050.
VE-High Detail	.2240, n.s
Mockup	0.0150, n.s.

To further understand the nature of these correlations, we plotted training transfer performance for the VE and hardware trained students against their visualization aptitude scores. The plot for the students trained in the low-detail VE and on the hardware mockup is shown in Figure 13. As can be seen in this figure, VE-trained students with high visualization aptitudes achieved similar combined performance scores to those trained with the hardware mockup. For those with low aptitude scores, however, combined performance was much lower than any hardware-trained student. Plots of the other groups using other performance measures produced similar results.

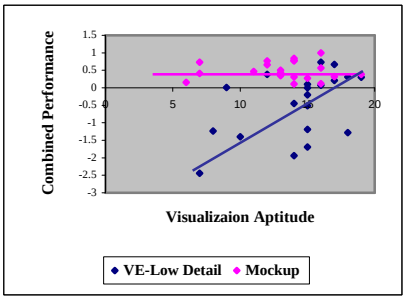


Figure 13. Plots of visualization aptitude by knowledge test score for hardware and immersive VE trained students.

The most straightforward interpretation of these findings is that the visualization aptitude mediates the effectiveness of VE-based training. That is, individuals with high visualization aptitudes can be expected to learn effectively from VE-based training by using their capacity to process visual/spatial information to gain the knowledge needed to perform on the job. On the other hand, students with low visualization aptitude scores can be expected to gain little from this type of instruction.

To further understand the mediating effect of the visualization aptitude, we evaluated the possibility that the task that we used in these studies put atypical demands on the visualization aptitude for VE-based training. That is, the task was selected because of the physical and visual obstructions involved. As such, it provided a critical test of the assumption that VE technology could be used to provide visual detail in training that would not be present in the real world. It was, however, not typical of maintenance procedures, which are generally performed in the visual field. Thus, a crucial question was whether VE-based training provided a more consistent effect when used to train more typical maintenance actions. In order to test this additional hypothesis, we modified the original fuel vale task so that hidden parts and structures were located now within the visual field.

Modified Fuel Valve Study

The maintenance-training task used in this study was a modified version of the removal and replacement of an aircraft fuel valve. The only difference between the original version of the study and the modified version was the placement of the fuel valve. In the original task, the fuel valve is mounted within the compartment towards the back of the right side wall. The position of the valve limits visual access to fasteners that are removed during the task. This means that a large portion of the mockup task is done only by "feel". In the revised version of the task, the fuel valve was moved directly underneath the access door, thus improving visual access to all critical parts. This modification of the task enabled us to test whether the VE-based training, which is primarily a visual simulation of the task, more closely approximates the training received by the hardware mockup when the task is learned and performed in the visual field.

Results

Results of the modified fuel valve study were compared with those of the original study. Summary statistics for the mean combined performance data has been standardized for comparison purposes on the chart in Figure 14. This chart plots mean composite performance for mockup training as well as VE training, both for the original fuel

valve task and the revised task. In addition, the 5th and 95th percentile scores are plotted to reflect variance.

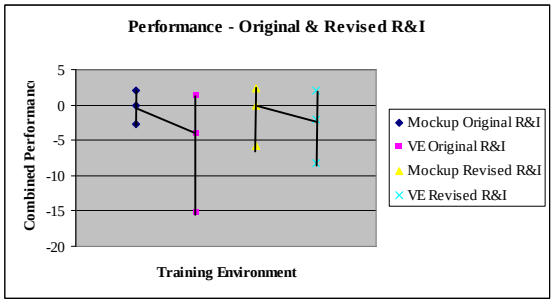


Figure 14. Mean performance and variation for original and modified VE training.

This chart illustrates that the VE-based training performance more closely matches that of the mockup-based training in our revised fuel valve task. While there are obvious differences in mean scores and variance between mockup and VE-based training in the original "tactile" task, the differences between these two types of training are much less pronounced for the revised "visual" task. Mean performance scores for the original task drop when comparing the VE to the mockup training. The 5th percentile for the VE-based training shows that many of the students did not learn effectively within the VE. This drop in learning correlates with the visualization aptitude—those with low visualization aptitude did not learn within the VE. For the revised "visual" task, the performance decrement from the mockup to the VE is not as pronounced, and the variance for the two groups is nearly identical. This pattern of results suggests that, for tasks performed in the visual field, VE-based training is much more effective, and comparable to that of hardware mockup training.

Inlet Ice Detection Fault Diagnosis Task

The second study on VE task selection investigated changing the nature of the VE training task. In this study, individuals were trained to perform a fault isolation task for the Inlet Ice Detection system. This task required the trainee to verify that a fault exists using fault codes within the cockpit, and then to diagnose the cause of the fault by performing continuity checks between connectors in two separate maintenance bays. Once the fault was isolated, the trainee needed to replace a part and then verify that the fault was corrected. This problem differed from the initial fuel valve remove-and-replace task because it involved multiple equipment bays and detailed continuity testing, but the tasking requirements were primarily visual. There were no complex motor skills involved in this task.

Aside from different task content, the experimental procedure was identical to the previous studies. Trainees were randomly assigned to either the mockup or VE-based training condition. Participants completed one pass through the training. After the training was finished, participants completed a written test of their task knowledge, and then completed the training transfer task on the full-scale hardware mockup. They were scored on how well they performed the task based upon measures of task performance time and errors. The score on the knowledge test, as well as the measures of task time and errors on the training transfer task were combined into a composite performance measure.

Results

Data from this study are plotted in Figure 15, along with the data discussed earlier from the fuel valve studies. The pattern of results from this study, a primarily visual task, are very similar to those observed from the modified fuel valve task (also primarily visual). The variance within the VE-based training group is very comparable to that of the mockup-training group for the troubleshooting task, and the mean scores for the VE-based training are comparable to those of the mockup training scores.

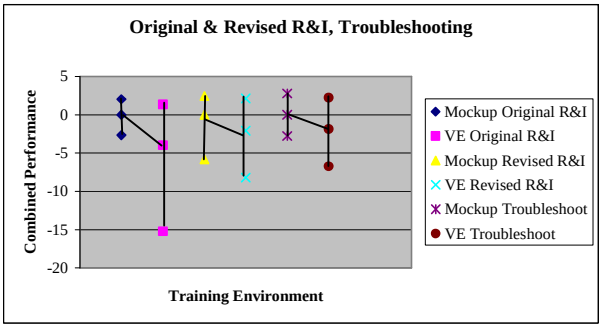


Figure 15. Mean performance and variation for original and modified VE task as well as troubleshooting task.

Collectively, this pattern of results suggests that VE-based training is effective for maintenance tasks that are learned and performed in the visual field, as opposed to those with a significant tactile component.

SUMMARY AND CONCLUSIONS

Based upon the results of numerous years of research into VE applications for maintenance training, we are able to draw a number of conclusions and summarize this work into empirically driven guidelines. First, for the bulk of maintenance activities for which visual feedback is available, VE-based training may provide learning and transfer performance that is largely equivalent to hardware-based training. Such a conclusion assumes, of course, that the visual cues provided by the VE are of sufficient resolution and detail to be similar to those available in the real world. Compared to hardware-based training, however, the VE would provide training at significantly reduced development and update costs and with significantly increased student throughput. Additional

research is underway to further refine these costing assumptions.

VE implementations that achieved high rates of training effectiveness included high-detail graphics, interactivity of the student with the training environment, incorporation of visual cues to give size perspective to the environment, and color-coded cues within the VE that match the actual environment. In our case, use of transparent aircraft skins provided the trainee with additional perspective as to where the aircraft parts were located beneath the door. Other potential VE implementations that may help to further engage the student in the training exercise include requiring the student to select the relevant tools from a "virtual toolbox" or menu, and allowing the student to adjust camera views, therefore providing more opportunities for student freeplay and interaction. We will be evaluating some of these factors in later studies. Our research failed to find effective training transfer in VEs where trainee interaction was not required, where perspective line drawings replaced higher-fidelity texture gradients, shading, and color of the traditional VE interface. Shifting camera views frequently also caused difficulties for some.

In observing which tasks are best suited for VE training, we found that tasks that are primarily visual in nature are the best match for the interactive desktop trainer. We saw training effectiveness improve dramatically when a relatively simple change was made, i.e., moving the replaceable part from being visually inaccessible to being in view. Although the physical components of removing fasteners and parts is still simulated within the desktop trainer, as long as the parts were within the visual field, the visualization demands from complex mental rotations and camera view adjustments are reduced. Consequently, training effectiveness is increased, even for those who otherwise have a difficult time processing complex spatial relationships. This seems true regardless of whether the task is procedural or diagnostic in nature.

For the minority of maintenance tasks where visual cues during performance are limited, the practice of supplementing these cues with enhanced visual feedback in a VE is questionable. To do so appears to draw on aptitudes in VE training that are not used in the real world. As a consequence, learning and skill acquisition from the VE do not consistently transfer, but rather, depend on the trainee's capacity to interpret and generalize from visual data to a non-visual task. At a minimum, VE training developers should verify that post VE-training performance is similar, both in level and in variability to that following hardware training, prior to fielding VE-based training that supplements the visual cues that are routinely available.

VMT technology enables several additional capabilities that are not possible for their hardware-based counterparts. For example, VMTs are deliverable as web-available, Advanced Distributed Learning (ADL™) conformant content, making them readily distributable, portable, and reusable. Similarly, using advanced capabilities of the ADL Sharable Content Object Reference Model (SCORM®) 2004, VMTs could be used to provide adaptive learning. Under this capability, individualized training is provided, based on trainee performance data, with initial investigations (e.g., Perrin, Dargue, & Banks, 2003) suggesting substantial increases in training effectiveness. In short, VMTs enable a host of additional capabilities, simply from being software rather than hardware based.

Additional research efforts will focus on two fronts: a) validating the effectiveness of these additional training interventions within the VE training environment; and b) investigating these training concepts using VE-specific hardware (e.g., advanced 3-D display technologies). It is this combination of innovation, coupled with empirically validated guidelines for implementation, that holds the most promise for optimized integration of VE technology within the training environment.

REFERENCES

Barnett, B., & Perrin, B. (2002). Leveling the field: Reducing differences in learning from virtual environments. *Proceedings of the Interservice/Industry Training, Simulation & Education 2002 Conference*, Orlando, FL.

Cook, T.D., Campbell, D.T., & Peracchio, L. (1990). Quasi-experimentation. In M.D. Dunnette & L.M. Hough (Eds.), *Handbook of industrial and organizational psychology*. Palo Alto, CA: Consulting Psychologists Press.

Ekstrom, R.B., French, J.W., Harman, H.H. & Dermen, D. (1995). *Manual for Kit of Factor-Referenced Cognitive Tests*, Princeton, NJ: Educational Testing Service.

Garner W. R. (1974) *The Processing of Information and Structure*. Potomac MD: Erlbaum.

Perrin, B., & Barnett, B. (2001). Virtual environments for training: Not just for pilots anymore? *Paper presented at the Image 2001 Conference, Scottsdale, AZ.*

Perrin, B.M., & Barnett, B.J. (April, 2002). *Is VE-based training for everyone? The mediating effect of visualization in training transfer*. Paper presented at ITEC 2002, Lille, France.

Perrin, B.M., Dargue, B.D., & Banks, F.Z. (2003). Dynamically adapting content delivery: An effectiveness study and lessons learned. *Proceedings of the Interservice/Industry Training, Simulation & Education 2003 Conference*, Orlando, FL.