This is a manuscript of an article from IEEE Transactions on Visualization and Computer Graphics (2014): 1, doi:10.1109/TVCG.2015.2393871.

Virtual Training: Learning Transfer of Assembly Tasks

Patrick Carlson, Anicia Peters, *Member, IEEE,* Stephen B. Gilbert, *Member, IEEE,* Judy M. Vance, *Member, IEEE,* and Andy Luse, *Member, IEEE*

Abstract—In training assembly workers in a factory, there are often barriers such as cost and lost productivity due to shutdown. The use of virtual reality (VR) training has the potential to reduce these costs. This research compares virtual bimanual haptic training versus traditional physical training and the effectiveness for learning transfer. In a mixed experimental design, participants were assigned to either virtual or physical training and trained by assembling a wooden burr puzzle as many times as possible during a twenty minute time period. After training, participants were tested using the physical puzzle and were retested again after two weeks. All participants were trained using brightly colored puzzle pieces. To examine the effect of color, testing involved the assembly of colored physical parts and natural wood colored physical pieces. Spatial ability as measured using a mental rotation test, was shown to correlate with the number of assemblies they were able to complete in the training. While physical training outperformed virtual training, after two weeks the virtually trained participants actually improved their test assembly times. The results suggest that the color of the puzzle pieces helped the virtually trained participants in remembering the assembly process.

Index Terms—learning transfer, haptics, virtual reality, assembly, training

1 Introduction

Training assembly workers to perform tasks in factory settings is of critical importance due to the cost, time, effectiveness, and safety of the process. Training is often done on the assembly line, sometimes causing a loss of productivity. In addition, the assembly process may be such that the unit cost of the parts makes training expensive. The cost for forgetting procedural tasks is high in production environments, where workers are trained and expected to retain the skills learned after a period of time without relearning [1], [2]. Providing training using virtual reality (VR) hardware and technology can offset these issues by ensuring a safe and potentially faster training environment. In order to test the efficacy and speed of virtual training, we devised a user study examining traditional physical training compared to virtual training. The base idea of comparing the training environments comes from a previous user study [3] where a six piece burr puzzle was used for the assembly task. In a between groups design, training was done either physically or virtually followed by testing with real puzzle pieces. Participants were timed assembling the puzzle. Spatial ability, technical experience, videogame experience, colorblindness, as well as other potential influencing factors were measured. Our previous research showed that virtually trained participants performed two times faster in the testing phase than those who were physically trained, despite longer training times [3]. We not only wanted to build a more robust study to attempt to replicate these results but also examine the role of color as a contributing factor, since it is commonly used to help distinguish parts in the virtual environments. This study represents some of the first work to compare learning gains from virtual training with physical training using a haptic device and data glove. The intent of this research is to inform the design of high-fidelity training within virtual environments.

- *P. Carlson is a PhD Human-Computer Interaction (HCI) graduate student at Iowa State University.*
E-mail: carlsonp@iastate.edu
- *A. Peters was a PhD Human-Computer Interaction (HCI) graduate student at Iowa State University when this work was completed.*
E-mail: anicia.peters@gmail.com
- *S. Gilbert is an assistant professor in the Industrial and Manufacturing Systems Engineering department at Iowa State University.*
E-mail: gilbert@iastate.edu
- *J. Vance is a professor in the Mechanical Engineering department at Iowa State University.*
E-mail: jmvance@iastate.edu
- *A. Luse is an assistant professor of Management Science and Information Systems at Oklahoma State University.*
E-mail: andyluse@okstate.edu

2 Literature Review / Related Work

Transfer of learning was first studied as transfer of practice when Thorndike and Woodworth [4] studied how learning one task can be applied to another task. Transfer of learning or learning transfer is the amount of knowledge gained during training that can be applied to a new task [5]. According to Gick and Holyoak [6] learning transfer can occur only when there is task similarity or domain knowledge and the person is able to perceive the similarity. A task that is structurally and surface similar will result in positive learning transfer whereas the opposite is true when there is structural and surface dissimilarity [5]. Allen et al. [7] distinguished between physical and functional fidelity when investigating learning transfer of simulators. In the theory of situated cognition, Greenco et al. [8] posit that transfer occurs when similar affordances and constraints are present in two dissimilar environments like in virtual and real world environments. Learning transfer and training recall will be successful according to the constructivism approach as long as the environment, goals, and cognitive processes activated are the same as those during training. Learning transfer is associated with skill acquisition, the learning curve, and skill retention. We will discuss these in turn. Our goal is to examine learning transfer under different training conditions.

2.1 Learning Transfer

Adams et al. [9] discussed three components of human training as being cognitive, perceptual, and motor demands. Users form an internal mental model and develop strategies for task completion. Any mistakes are corrected and noted and motor demands involve the handling, orientation, and connecting of physical pieces.

Fitts three-stage skill acquisition process is well-known [10]. The first stage is the cognitive stage where the learner identifies how something works. The second associative stage is where the learner mainly corrects errors made during the cognitive stage. Finally, the autonomous stage is where the learner shows a gradual improvement indefinitely [11]. However, the plateauing learning theory by Conway and Schultz [12] states that there is a plateauing or stabilization of improvement.

Anderson [13] expanded Fitts' staged theory of skill acquisition by adding a declarative stage, a knowledge compilation stage, and a procedural stage. Errors and verbal mediation are common in the declarative stage, whereas verbalizations disappear, errors are corrected, and speed increases during the knowledge compilation stage. The procedural stage is identified by selecting a few alternative solutions. Fitts' theory was used over Anderson because of the simplicity of the puzzle assembly task and the fact that there was no verbalization or alternate puzzle solutions.

The learning curve was first discovered by Wright [14] when investigating factors influencing the cost of airplanes. The basic factor attributed to the curve is that for every doubling of units produced, the labor rate decreased by a constant percentage. Since then, manufacturers have taken a keen interest in the cost benefits associated with predicting their employees' learning curve [2]. Varieties of the learning curves are the log-linear model, the plateau model, the DeJong model, the Stanford-B model, and the S-model [2]. Although there are several learning theories presented in the literature, we were particularly interested in the plateauing model of Conway and Schultz [12] (pg. 37) as we predict that the virtual training will have a different learning curve and stabilization point (plateauing) than physical training.

Skill retention strongly correlates with skilled performance, but the amount of skill retention is dependent on task and situational factors [15]. Arthur et al. [15] lists among others the degree of overlearning, conditions of retrieval, and motivation as moderators for skill retention. Overlearning leads to automaticity and should reduce the amount of skill decay [15]. Lance et al. [16] has shown that overlearning can actually increase skill decay. They define overlearning as a learning saturation point when any additional learning causes a drop in performance or skills decay. Hall et al. [17] found in their study that extended practice time did not increase skill retention. Based on these results, overlearning and skill decay should be evident in both physical and virtual testing.

2.2 Learning Transfer in a Virtual Environment

Virtual training has proven to be beneficial in a number of areas such as flight simulator and military training [9], medical training [18], [19], sports and rowing training [20], and route information and mapping tasks [21].

Our focus is assembly tasks, which is defined by Boud et al. [11] as, the manipulation and joining of parts to form a whole. Although virtual training of an assembly task have been studied for a long time, there are yet no proven benefits over physical environment training [21]. The literature shows mixed results in numerous studies as examined by Hamblin as part of dissertation work [5]. In one study of an assembly task in a virtual environment, they found that the real world training was more effective than virtual training [5]. Adams et al. [9] performed a study where a Lego biplane model was constructed to investigate learning transfer from a virtual environment with the use of haptic force-feedback and concluded that haptic feedback was necessary for more efficient learning transfer. Hall and Horwitz [22] in a follow-up study from an initial study in 1998 investigated procedural knowledge retention after a period of time when learning took place in a virtual environment compared to a conventional 2D computer environment and reported no significant differences between the groups. Rodriguez et al. examined learning transfer of procedural tasks within a multi-modal virtual environment by building a model lego plane [23] while Gerbaud et al. created an entire software platform for teaching procedural tasks using VR [24]. The rich fidelity of VR environments provide a wide variety of learning affordances [25].

In order to study the learning transfer effect of an assembly task within a virtual environment, it is necessary to consider and design a virtual environment that will optimize learning, but not necessarily merely replicate the physical environment. Moeser [26] showed that, "people can be in a real environment for more than a year without necessarily acquiring a survey representation (configurable knowledge) of it, but it remains merely route representation". We need to consider that virtual environments also differ from physical environments in that mostly cognitive learning takes place in the former and motor skills in the latter [5]. In a study designed to test learning of a route through an office building [27], it was shown that virtual training can be just as effective, but Piller and Sebrechts [28] showed that making the walls along the route transparent actually improved the memory of the virtual participants and they performed better in the real world when tested. We considered these factors in our design of the virtual environment. During the testing phase, we identified the glove was occluding objects and participants could not see the angle of rotation. Participants moved their virtual hand away from the object in order to view the angle of rotation before rotating it further. Based on this, we made the virtual representation of the glove transparent which allowed participants to see the part that they were handling. This was important to ensure occlusion did not negatively affect learning.

Recent research in comparing different visual feedback types for virtual grasping showed that the see-through method that we used was one of the best methods for translational performance for small objects [29]. While not rated as highly in other areas such as subjective preference, accuracy is of vital importance in the assembly process.

Selecting a puzzle for an assembly task has been successfully used in the past by Shuralyov and Stuerzlinger [30]. In a small study, participants were asked to assemble a 3D puzzle using a mouse and keyboard on a desktop LCD. Task performance time for the assembly was acceptable, however, it was noted that participants spent a considerable amount of time rotating the pieces. Ritter et al. [31] used a 3D puzzle since participants would already know how to interact with the puzzles, and shapes provide cues of what the structure of the puzzle is.

While gender differences were observed when using paper and pencil measures, no such differences were observed in the virtual environment tests [32] where 3D drawings were used instead of the conventional 2D drawings. Parsons et al. [32] speculated that the strategies employed by males and females were different which yielded differences in performance. Parsons et al. [32], Larson et al. [33], and Rizzo and Buckwalter [34], tested gender differences for mental rotation in virtual environments and found that gender difference for mental rotation vanishes in virtual environments. In our studies, we considered that spatial ability might be a moderator of performance, so we measured this beforehand with a 2D Vandenberg and Kuse mental rotation test [35].

Fig. 1. Six Piece Wooden Physical Burr Puzzle

Generally individual differences such as gender, age, videogame experience, prior technical computer literacy, and computer efficacy affect learning transfer in virtual environments [36]. We will consider these factors in our study.

The primary motivation for the work presented here came from a previous study conducted in 2011 that examined whether training in a virtual environment was as effective as training in a physical environment [3]. Participants were randomly assigned to a virtual environment (VE) or a physical environment (PE) to assemble a six-piece burr puzzle (Figure 1). A burr puzzle is wooden cube-shaped interlocking 3D puzzle. The training instruction sheet used color coded individual pieces for easy identification. Participants were given unlimited time to complete the assembly task and were asked to assemble it two times with the aid of the instruction sheet. After training, performance was tested on the assembly with the physical puzzle without any instructions and without color indicators on the physical puzzle pieces. The results showed that the VE participants took three times as long as the PE participants to complete the training, but the VE participants completed the puzzle in the testing phase almost two times faster than the PE group. There was a large variance in the time the participants took for the assembly training. We wanted to examine both the effects that the unlimited training time and the use of the color indicators had on the participant's puzzle assembly performance in the study described here.

3 Methods

Based on the open questions in the research on transfer of training in VEs outlined above, we will study learning transfer (skill acquisition, learning curve, skill retention/decay) in the virtual environment and in the physical environment using the same six-piece burr puzzle as a proxy for assembly parts. The six piece burr puzzle provided a familiar concept, yet a sufficiently complex model in which the participant had to follow the instructions in a specific sequence in

Fig. 2. 5DT Glove (left hand) and Phantom Omni® (right hand) in virtual training

order to gain procedural knowledge to put the puzzle together.

3.1 Hardware and Software Environment

We created a multi-sensory environment that incorporated visual and haptic force-feedback (no auditory feedback) to make the interaction between the user and the puzzle pieces as real as possible. Rodriguez et al. discuss a wide variety of benefits using multimodal virtual environments [23].

Studies of asymmetric and symmetric bimanual tasks have resulted in a series of insights into the way humans use both hands when completing various tasks. According to Hinckley et al. [37], bimanual interaction is optimal when each hand assumes its most effective role. The most common bimanual interaction is one in which the non-dominant hand is responsible for gross motor movements while the dominant hand performs more fine motor positioning [38], [39]. Supporting research indicates that the non-dominant hand is generally used for lower-frequency and high-amplitude movements and the dominant hand is used for higher-frequency and lower-amplitude movements [40]. These results motivated our choice of devices with the 5DT glove in the non-dominant hand being useful for broad range motion and positioning while the Phantom Omni® with haptics was used in the dominant hand for fine motor control of the individual pieces.

3.1.1 Hardware/Software

The virtual training utilized stereoscopic glasses, a rear projected desktop stereo image (1280x720), an Intersense IS-900 hybrid ultrasonic and inertial tracking system for tracking the head position, a 5DT Data Glove 5 Ultra, a Polhemus Patriot magnetic tracker for tracking the glove, and a Phantom Omni® with haptic force-feedback. The Phantom Omni® and 5DT glove can be seen in Figure 2.

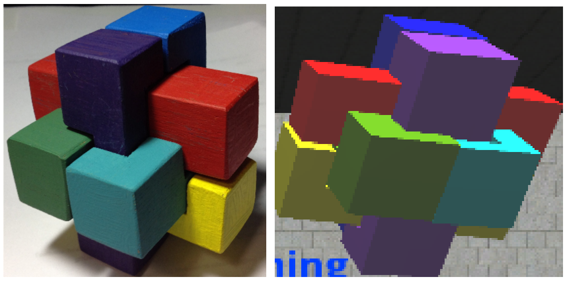

Fig. 3. Physical (left) and virtual (right) burr puzzles

The application was developed using SPARTA (Scriptable Platform for Advanced Research in Teaching and Assembly) [41]. SPARTA combines VR Juggler [42] for stereoscopic rendering, OpenSceneGraph for graphics, Voxmap PointShell (VPS) [43] for physics calculations, and VR JuggLua [44] for easy scripting and content creation. It supports multiple input and output devices including position trackers, stereo glasses, stereo projection systems, gloves and haptic devices. The software was developed by researchers at the Virtual Reality Applications Center at Iowa State University.

3.1.2 Participant Interaction in VR

Participants were seated in front of a glass display and haptic device. They put on stereo glasses and a glove which can be seen in Figure 2. They held the stylus of the haptic device in their right hand and wore the glove in their left hand. They could then manipulate the 3D environment by grabbing and manipulating the virtual puzzle pieces with either the glove or haptic device. The pieces could be grouped together and moved as a single unit (subassembly) by toggling a virtual button in the environment. While in this mode, the button on the haptic device was the only way to add pieces to the subassembly. Holding the button on the haptic device allowed the participants to pick up and move the pieces. The haptic device provided force-feedback to make the participant aware that the pieces were colliding.

3.2 Experimental Design

To assist participants in following the instructions during assembly training time, we colored each piece of the puzzle a distinct color and used those same colors on the instruction sheet. We used the physical burr puzzle to train participants in the physical group and a virtual desktop training environment for the virtual group (see Figure 3). Each of the six parts has the same general shape with slight variations. Because of this and the general difficulty in assembling the puzzle we felt using the puzzle was comparable to more manufacturing oriented assembly scenarios. The use of the puzzle also allowed us to recruit participants from non-engineering fields.

TABLE 1
Independent Variables

Independent Variable	Classification	Group
Training Condition	Between	Physical, Virtual
Color Order	Between	Color First, Wood First
Test Order	Within	Initial Test, Retention Test

From the results of the first study, we decided to investigate further what caused the differences in learning transfer between the two groups. We redesigned the original study to control for the following variables. Training times were held constant between virtual and physical training. Participants went through an equipment familiarization video.

In order to reduce cognitive load and utilize the benefits of the virtual environment, a few features were changed. First, we replaced the physical manual button participants had to press to toggle subassembly mode with a virtual button. This saved time and prevented the participants from having to look away from the virtual interface. Second, the virtual representation of the glove was made transparent to ensure participants could see the part that they were manipulating. This prevented the representation of the glove from occluding the pieces.

3.2.1 Study Design

For this study, a mixed model design in a controlled lab experiment was chosen. The dependent variable was task performance as measured by the amount of time to assemble the puzzle. We wanted to see during testing what influence the coloring had on performance. Previous research examining the role of color on mental rotation tasks identified that individuals with strong mental rotation abilities do not use color identifiers on the piece itself during the mental rotation while individuals with poor mental rotation abilities do use color identifiers [45]. Participants were tested on both the physical color pieces that matched the training and plain wooden pieces that were not colored. The independent variables were training environment (Physical vs. Virtual), and color order of the first test (first color, then wood, or first wood then color). Both of these were designed as between participants tests. Another independent variable was the testing time. An initial test immediately after the training period as well as a follow-up test two weeks later was performed to assess the learning transfer. This was a within participants test. The breakdown of independent variables can be seen in Table 1. A pilot study was conducted with eight participants who performed equally well in both the virtual and physical environments, but they were almost exclusively engineering students with high spatial abilities.

3.2.2 Procedure

A broad overview of the procedure can be seen in Figure 4. Using blocking, participants were randomly assigned to groups with consideration taken to roughly distribute men and women between the groups. This was done to account for similarities among groups of subjects, thereby helping to alleviate unequal variances which can occur when doing completely random sampling [46].

After an informed consent was obtained, participants were asked to complete a brief questionnaire containing demographic questions about their computer usage and expertise, videogame playing, number of engineering courses completed, and educational background. The Ishihara color blindness test [47] was administered to control for variances due to color blindness.

Participants were given a timed 10 minute redrawn Vandenberg and Kuse Mental Rotation Test (MRT) [35]. The test measures spatial ability and presents the participant with 20 questions that require matching a target shape to the corresponding rotated shape. The test also includes eight questions which asked participants what type of strategy they utilized for the mental rotation task. This test was selected due to its reliability shown in a previous study [35].

Next, all participants were shown an equipment demonstration video which demonstrated the equipment to be used in the virtual environment and how to carry out subassembly grouping. Subassembly grouping allowed the participants to pick up, move, and rotate multiple objects as a single group. The video showed a user manipulating small virtual cubes/boxes to demonstrate assembly which were different from the burr puzzle pieces. We considered the inclusion of a video demonstrating the equipment to be used as beneficial for the virtual assembly task [11]. Participants in the physical condition also watched this video since they were not informed beforehand which condition they were assigned.

The virtual training participants were then given eight minutes to familiarize themselves with the equipment and VR setup by performing a box-stacking task that required use of the subassembly feature. Participants were free to request assistance and ask questions during this period. Participants in the physical training group were given an online crossword puzzle as a filler task. The crossword puzzle filler task was deemed suitable as per past studies in virtual environments [34].

Next came the assembly training period. All participants were given 20 minutes training to assemble the six-piece burr puzzle according to an instruction sheet placed in front of them. The puzzle pieces on the instruction sheet were colored to match the colors of the physical pieces and the virtual pieces. Participants were asked to assemble the puzzle as many times as possible and were timed on each successful

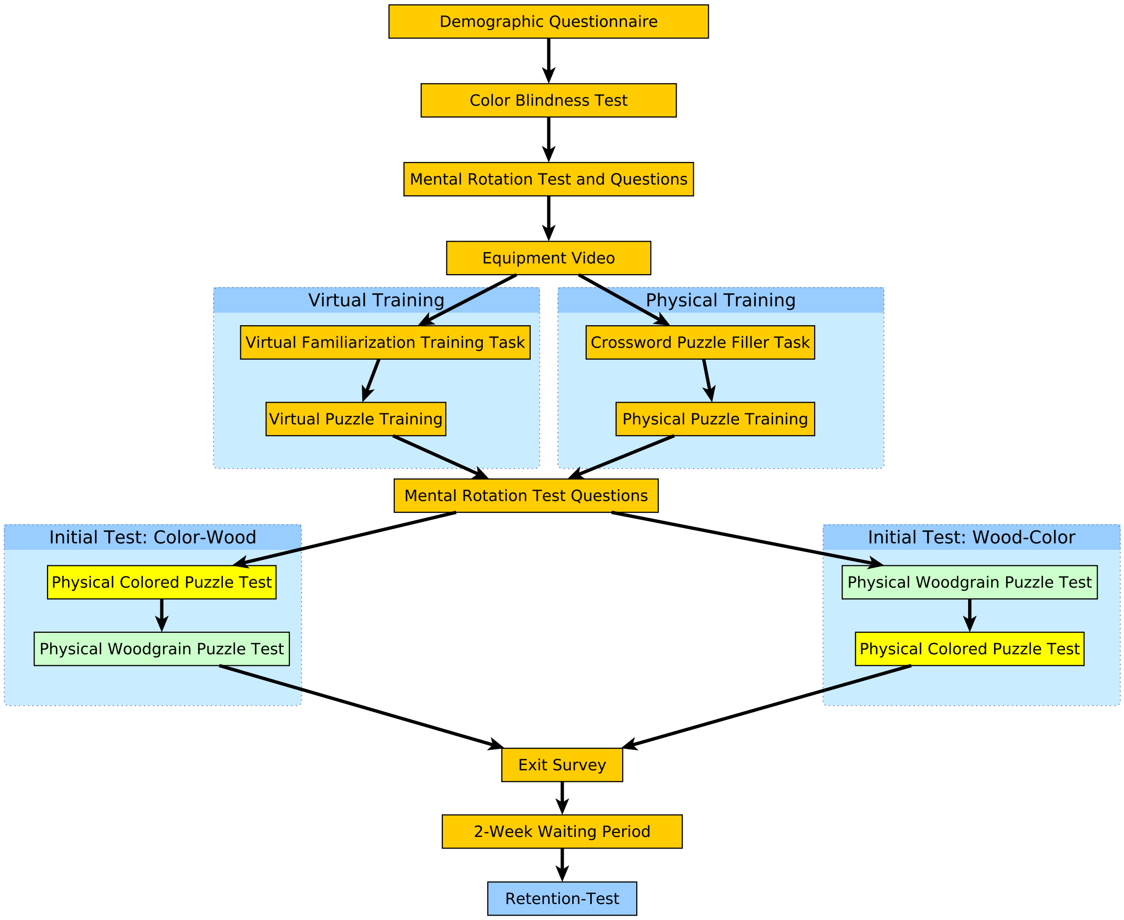

Fig. 4. Overview of study design and random assignment

completion. In both virtual and physical training, the puzzle pieces were randomly positioned and oriented after each successful assembly in order to mimic the physical environment. The software tracked the completion times of the virtual environment and a second researcher recorded times in the physical training environment. The times were recorded in order to measure the learning curves and learning saturation point (stabilization plateau) of all participants.

After the training phase, participants were given a set of eight Vandenberg and Kuse MRT questions about the strategies they used during training. These questions also served as a distractor task in order to eliminate any recency effect that might occur. The recency effect refers to the potential that a participant's performance is more highly influenced by a recent task than by a task performed less recently [48]. An initial test was administered by asking all participants to assemble the physical puzzle without instructions. There were two sets of physical puzzles used in the testing: one had all pieces colored and one had all the pieces the natural wooden color. Depending on whether the participant was assigned to a color-first or wood-first treatment condition, the participant was either given a colored puzzle or a wood grain puzzle to assemble first, followed by the other puzzle. Participants were instructed to assemble each puzzle three times within a ten-minute period.

An exit survey was administered which asked participants what recall strategy they used, the ease of use and helpfulness of the training environment, difficulty of the task, and overall realism of the training environment.

After two weeks, participants returned and performed a retention test which was the same as the initial test of assembling the physical burr puzzle. The same color-first or wood-first treatment was used for each participant as in the initial test. Participants were instructed to assemble each puzzle three times within a five-minute period. During the initial test, both groups of participants encountered the physical wooden puzzle for the first time where both color and the instructions were removed. For the second test after two weeks, both groups had equal experience levels with the physical puzzle. To account for learning effects, we factored in additional time for the initial test (10 minutes) as compared to the second test after two weeks (5 minutes).

TABLE 2
Number of Participants by Condition

	Color First Testing	Wood First Testing
Physical Training	Female: 4, Male: 13	Female: 6, Male: 10
Virtual Training	Female: 6, Male: 10	Female: 6, Male: 8

4 RESULTS

4.1 Participants

Sixty-three participants (22 females and 41 males) completed the study. Some participants received class credits as compensation by enrolling in the study through the psychology or marketing departments. Sixteen participants received 5 USD compensation. The participants were mostly undergraduates with ages ranging from 18 to 31 years old (M = 21.62, SD = 2.64). Five participants tested positive for being colorblind. The breakdown of participants in each condition can be seen in Table 2. Given that not all participants finished all tests, we wanted to report the most conservative test possible for each separate test so instead of using mean imputation, we used pairwise deletion to remove the individual that was missing the value when appropriate for the following tests.[1] All results were analyzed at a confidence level of 95%. The differences between genders on different factors can be seen in Table 3.

There were no significant differences in the MRT score $t(61) = -0.52$, $p = 0.59$ between the group of participants performing the virtual training and the group of participants performing the physical training. This shows that the relative spatial abilities between the groups is roughly equal and should not overly influence the testing times. Even so, MRT was included as a covariate in later tests. The number of engineering courses taken by participants in each of the virtual and physical training groups was also not significantly different $t(61) = -0.96$, $p = 0.34$. There was a low correlation between video game playing and MRT score, $r(60) = 0.28$, $p = 0.02$. There was no significant correlation between video game playing and the number of puzzles assembled during training, $r(60) = 0.009$, $p = 0.94$.

4.2 Training Results

In comparing the number of completed puzzles during training, those who were trained physically completed significantly more puzzles than those who were trained virtually, $t(61) = 10.01$, $p < .001$. In addition, those who were trained physically needed significantly less time to assemble the puzzles during training, $t(48) = -6.73$, $p < .001$. The correlation between the number of assembled puzzles done in training and participant MRT score were weakly correlated, $r(61) = 0.34$, $p = 0.005$. Participants who were trained with the physical puzzle, $r(31) = 0.79$, $p < 0.001$, showed a significantly stronger correlation, $z = 2.704$, $p = 0.006$ [49], [50], than those who were trained with the virtual puzzle, $r(28) = 0.34$, $p = 0.05$.

4.3 Test Results

A mixed-factorial ANOVA analysis was used to test the omnibus model using SAS proc mixed. Only participants who finished all levels of the within-subjects variables were included in the analysis. This ensured balance between the groupings. A significant 3-way interaction exists (see Figure 5) between the three independent variables of testing session (initial vs. retention-test), color order (color first or wood first), and training environment (PE vs. VE) on the dependent variable of test time, $t(28) = 2.26$, $p = 0.03$.[2,3] To better understand the drivers of the interaction, simple 2-way interactions were investigated in the context of training environment (PE and VE respectively). The results show that the interaction of testing session and color order for those who were trained physically (the two solid lines in the figure) is not significant $t(28) = -0.93$, $p = 0.36$, with a simple main effect showing that time will significantly increase from initial to retention-test between 23 and 66 seconds regardless of color order when the individual is trained physically $t(28) = 4.16$, $p < 0.001$, CI=[23,66]. Conversely, the interaction of testing session and color order for those who were trained virtually (the two dashed lines in the figure) is significant $t(28) = -3.47$, $p = 0.002$. The simple-simple main effect shows that time will significantly increase from initial to retention-test between 15 and 96 seconds when tested with wood first $t(28) = 2.82$, $p = 0.009$, CI=[15,96], but will significantly decrease from initial to retention-test up to 90 seconds when tested with color first $t(28) = 2.13$, $p = 0.04$, CI=[2,90].[4] The breakdown of testing performance can be seen graphically in Figure 5 and numerically in Table 4.

Comparing the PE and the VE directly at the two respective testing times, we find that the initial times

1. We observed that a substantial number of participants in the virtual training condition did not finish the puzzle assembly as compared to the physical training participants. To further investigate this, we ran the same primary data analysis as the time data but used a dichotomous dependent variable of whether or not they completed the puzzle assembly test. The results show no three-way or two-way interactions, with only a main effect of training environment, $F(1, 58) = 16.31$, $p < .001$. However, the post-hoc power analysis shows this doesn't adversely affect the study results.

2. MRT was included as a covariate in the analysis to account for any confounding variance between subjects due to spatial ability.

3. Gender was tested as an independent variable but there were no significant interactions or main effects, thus we did not include it in the analysis.

4. Least square estimates are used to calculate the means, standard errors, and significance tests. Least square means allow for a more precise estimate of the means which are adjusted for the other effects in the model [51]. The SAS LSM estimator allows for the testing of all hypotheses/primary tests using contrast coefficients and produces an associated t-test to evaluate significance.

TABLE 3
Demographic Information: Differences by Gender

	Male Mean (Std. Dev.)	Female Mean (Std. Dev.)	*df, t*	*p*	Overall Mean (Std. Dev.)	Measurement Unit
Number of Engineering Courses Taken	2.80 (0.90)	2.95 (1.50)	$t(61)$ = -0.09	0.92	2.85 (6.22)	Range: 0-30
Computer Technical Expertise	3.45 (0.10)	3.23 (0.09)	$t(60)$ = -1.39	0.16	3.37 (0.60)	Likert (1=low to 5=expert)
Video Game Playing	3.18 (0.18)	1.86 (0.20)	$t(60)$ = -4.43	$<.001$	2.70 (1.27)	Likert (1=never to 5=play daily)
Mental Rotation Score	12.98 (0.92)	10.77 (1.34)	$t(61)$ = -1.38	0.17	12.21 (6.09)	Score Range: 0-20 (low-high)

for those who completed the training in the PE were significantly lower than those who completed the training in the VE whether they completed the testing using wood or color first. Stated more formally, the simple-simple main effect shows that time is significantly lower, by 12 to 68 seconds, for the PE with wood first $t(27) = -2.94$, $p = 0.007$, CI=[-12, -68] and anywhere from 54 all the way to 107 seconds lower in the PE with color first $t(25) = -6.28$, $p < 0.001$, CI=[-54, -107]. Similarly, time is also significantly lower at the $p < 0.1$ level for retention-test times for the PE with wood first $t(27) = -1.78$, $p = 0.086$, CI=[-2, -82]. Conversely, for those in the VE using color first, retention-test scores are not significantly different from those in the PE using color first $t(25) = -0.01$, $p = 0.99$.

A post-hoc power analysis was run to verify that the necessary power was achieved to enable the discovery of significant effects for all available tests. An alpha level of 0.05 was used along with a sample size input of 30 (conservative estimate of the lower of the two sample sizes). A four group power analysis was run (for the 2x2 combination of the two between-group variables) with two measurements (for each level of the within-subjects variable). A conservative correlation between measurements of 0.01 was used as this was the smallest correlation between initial and retention-test scores among the various combinations of groups. Effect sizes of Cohen's d were calculated for each difference in initial and retention-test least square means using the formula $d = \frac{|\bar{x}2-\bar{x}1|}{\sigma}$ [52] where σ is the pooled standard error of the differences between the pairs of least square means, or $\sigma = \frac{se}{\sqrt{\frac{1}{n1}+\frac{1}{n2}}}$, which Cohen indicates can also be used in a within-subjects design [53]. Using this equation, the smallest effect size between initial and retention-test scores was 0.55, which was used as a conservative estimate in the power analysis. To test for nonsphericity, the mixed model was run using both the Huynh-Feldt (hf) and unstructured (un) options for the covariance matrix calculation. The difference in the -2 Log Likelihood values of 13.3 with a difference in degrees of freedom of 5 distributed as a chi-square test rejects the null hypothesis of sphericity and implies some nonsphericity. Given there are only two within-subjects measurements, the sphericity correlation must be 1 based on the formula $(\frac{1}{numMeasurements-1})$. Using these inputs, we get a power of 0.93, which is well above the typical measure of 0.8 [54].

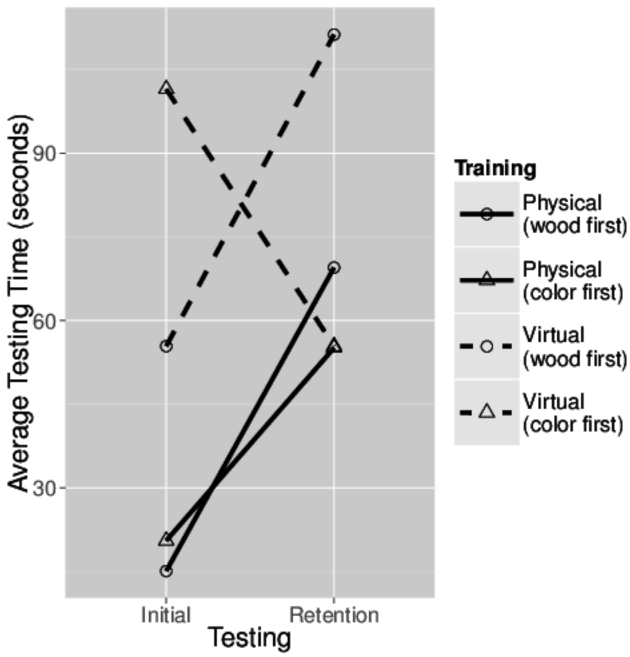

Fig. 5. Testing Performance Between Training Environments With Color Order

4.4 Learning Curve

Figure 6 shows the results for the learning curve of participants as measured during training. As the number of times they assembled the puzzle during the training period doubled, their puzzle assembly time was reduced by half. For those who were physically trained, the learning curve is quite short and they reached a plateau point quickly. The number of puzzles that the physically trained group assembled eventually reached a plateau of peak performance. Virtually trained participants initially exhibited worse performance than the physically trained participants, but also exhibited improving performance as the training progressed.

TABLE 4
Task Performance Time Results (in seconds)

		Initial Test	Retention Test	t value	p value	CI
Physical Training	Color First	15.09	69.49	3.37	0.002	[21, 87]
	Wood First	20.53	55.1	2.47	0.02	[6, 63]
Virtual Training	Color First	55.41	111.27	2.82	0.009	[15, 96]
	Wood First	101.48	55.41	-2.13	0.04	[2, 90]

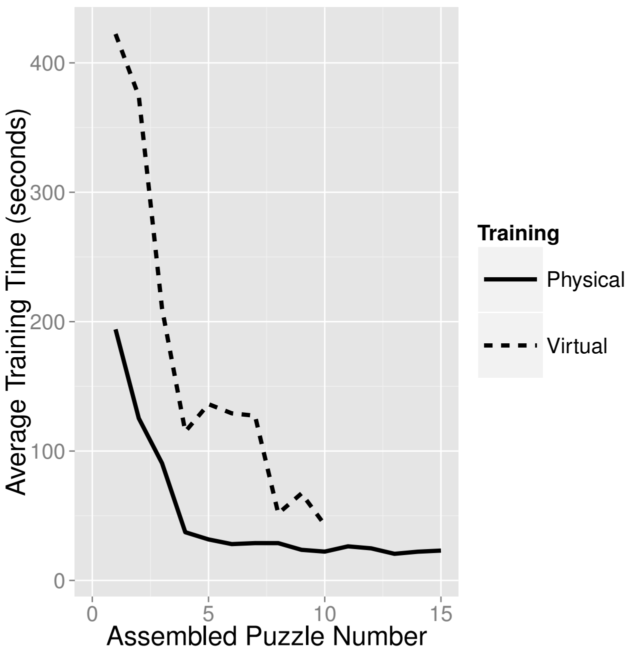

Fig. 6. Learning Curve Between Training Environments

4.5 Strategies

According to the self-reports on recall strategies used, the PE group reported to remember how the pieces were to be assembled, whereas the VE group relied primarily on color as a recall strategy. The PE group had the 10 fastest testing scores. These fastest PE participants said that they mostly used shape as their recall strategy. Comparing the 5 fastest mean test times for both physical and virtual, both shape and color were cited equally as a recall strategy.

After both the MRT spatial ability test and after the training, participants were asked a series of eight questions regarding the strategies that they employed for the task. The response data was categorical; therefore, a chi-square analysis was performed comparing the responses between the virtual and physical training groups. Two results for the questions asked after the training was finished were significant. One question asked how participants compared the matching of the target figure, χ^2 (2, N=63) = 7.29, p = 0.02. The range of approaches to solving the problem in terms of developing a specific approach as compared to using various approaches was also significantly different, χ^2 (2, N=63) = 13.30, p = 0.001. The results can be seen in Table 5. All other chi-square results were not significant.

At the end of the initial test, participants were asked questions regarding the difficulty, ease of use, and general realism. The overall results can be seen in Table 5.

Participants who were physically trained rated the difficulty as significantly easier than those who were virtually trained using a Likert scale (1=very difficult, 5=very easy), $t(60)$ = 4.42, $p < .001$. Participants who were physically trained rated the ease of use in assembling the parts in the training environment as significantly easier than those who were virtually trained using a Likert scale (1=very difficult, 7=very easy), $t(60)$ = 5.20, $p < .001$. There was no significant difference in how realistic the participants in each of the training conditions felt the training environment was depicted using a Likert scale (1=not at all realistic, 5=very realistic) which suggests the virtual training was fairly accurate. Interestingly, virtually trained participants rated the training environment for learning the assembly process using a Likert scale (1=not at all helpful, 5=very helpful) on average slightly higher (M = 4.03) than the physically trained group (M = 3.75). However, there was no significant difference between the two groups, $t(60)$ = -1.10, p = 0.27. When asked to rate how seriously they took performing the tasks using a Likert scale (1=very unseriously, 5=very seriously), there was no significant difference between the virtually trained versus the physically trained groups, $t(59)$ = -0.49, p = 0.62.

5 Discussion

The three-way interaction of testing session, color order, and training environment on the time needed to finish the virtual task offers interesting results. First, those individuals who completed the test in the PE performed significantly worse in the initial test as compared to the retention-test regardless of whether the individual performed the assembly using the wood or color puzzle first. This implies that the order of training with regard to wood or color order does not matter when the user is trained physically, with training times becoming significantly worse in the retention-test regardless. Conversely, the order of color and wood puzzles becomes a significant factor for those who are trained virtually. While

TABLE 5
Questions After Training

	Physical Training	Virtual Training	Test	p value
Comparison to Target After Training (Vandenberg and Kuse #5)	number of participants			
I always compared the options to the target figure.	16	24	χ^2 (2, N=63) = 7.29	p = 0.02
Once I found the matching puzzle piece, I compared the rest of the options to the match.	6	1		
I did a bit of both.	11	5		
Problem Solving Approach After Training (Vandenberg and Kuse #6)	number of participants			
I developed a specific approach to solve the problems.	23	9	χ^2 (2, N=63) = 13.30	p = 0.001
I tried various approaches to solve the problems.	4	16		
I did a bit of both.	6	5		
Exit Questionnaire	Mean (SD)			
Rate the difficulty of the tasks you completed. (1=very difficult, 5=very easy)	3.78 (1.03)	2.63 (0.99)	$t(60)$ = 4.42	$p < .001$
Rate the ease of use in assembling parts in the training environment. (1=very difficult, 7=very easy)	5.03 (1.61)	3.10 (1.26)	$t(60)$ = 5.20	$p < .001$
Rate how realistic you felt the training assembly environment was. (1=not at all realistic, 5=very realistic)	3.65 (0.86)	3.33 (0.95)	$t(60)$ = 1.39	p = 0.16
Rate how helpful the training environment was for learning the assembly process. (1=not at all helpful, 5=very helpful)	3.75 (1.07)	4.03 (0.92)	$t(60)$ = -1.10	p = 0.27
Rate how seriously you took these tasks. (1=very unseriously, 5=very seriously)	3.90 (0.58)	4.00 (0.88)	$t(59)$ = -0.49	p = 0.62

those trained virtually using wood first show the same degradation in speed in the retention-test, those trained virtually using color first show a significant improvement in speed in the retention-test. Furthermore, this improvement places these individuals at the same level of performance in the retention test as those trained physically. Given the advantages of virtual training with regards to time and money, this offers promise as a viable method for adequately training personnel that is equitable to a physical training environment.

This finding is important for assembly work where there are interruptions in production or time lapses between training and on-the-job performance, especially for procedural training tasks since it is costly to retrain workers. Our results are further consistent with Bailey's [1] finding that skill decay is not strictly related to the learning curve since the PE participants registered superior training times over the VE participants. However, there seems to be evidence of a small amount of overlearning once the automaticity phase and thus a learning curve plateau has been reached. This can actually lead to skill decay similar to what we witnessed for the physical environment [16], [15].

Although the training environment matched the testing environment and the long-term memory of the learned skill should have been activated during the retesting [55], this did not appear to be the case for the PE and it may have been the case for the VE.

5.1 Training and Learning Curves

Our observations during the training phase were consistent with Adams et al. [9] who defined three components of human training as cognitive, perceptual, and motor demands. We observed that participants build an internal model of tasks in their memory and put together a strategy to avoid common mistakes with easily confused pieces such as the teal and green pieces. In order to highlight the perceptual aspects, we used shape, color and sequence of the puzzle pieces. We also used three different sets of puzzles of both color and wood to avoid participants from learning physical friction points between unique puzzle pieces. For motor demands, we observed a dexterous manipulation of the pieces, how to handle them, move them, orient them and connect them in both the VE and PE. Motor demand benefits most from haptic feedback [9], but the misalignment of pieces due to tolerance and lack of snap-to functionality presumably contributed to slower performance for the VE participants.

Comparing the learning curves of both groups, the PE participants seemed to have reached a plateau or stabilization point after around the 6th puzzle completion and thus very little or no learning took place after that according to our results and also according to the learning curve theory of Conway and Schultz [12]. The VE participants appear to be moving towards this plateau during our training period, but probably did not reach it completely. The learning curve of the VE participants was quite dramatic showing a high increase in performance after only a few assemblies. A possible reason for the skill decay shown in the retest of the PE group is that when people reach a learning saturation point, there is an overlearning effect [9] where any additional learning may cause a drop in performance [16]. We noticed several times when performance for the PE participants dropped and then increased again, only to drop another time. This could also be attributed to fatigue since participants had to complete so many iterations of the puzzle assembly. We found no evidence that an increase in practice trials increased skill retention as observed for the PE participants who showed greater skill decay after two weeks. This finding is similar to Hall et al. [17]. We attribute this skill decay to the plateauing effect and its associated overlearning effect for the PE participants as discussed before [2].

We observed that the learning curve times were also consistent with the literature [2] with PE participants recording a decrease in times in every doubling of efforts until the plateau was reached. The learning curve stages observed for the PE participants were also consistent with Fitts' three-stage skill acquisition model comprising the cognitive, associative, and autonomous stages [10]. During the cognitive stage, we observed participants first followed the instructions during the assembly process until they reached a level of proficiency as is evident up to the fourth puzzle completion. Then they moved on to the associative stage by discarding the instructions and attempting it on their own. During the associative stage, they struggled quite a bit with the correct orientation for the teal piece as well as noticing the geometric difference between the green and purple pieces. Finally they reached the autonomous stage with only a minor improvement in time until finally saturation of performance and no performance gains could be seen. It is not evident from the data whether the VE group reached the associative stage since there was no jump in performance times as was evident for the PE group. The VE certainly never reached the autonomous stage.

5.2 Virtual Environment Training

We attribute longer training times in the VE group to the hardware and software environment. The cognitive load associated with learning the novel interface as well as attempting to learn the puzzle assembly was demanding. Puzzle parts in the VE environment did not have a snap-to function or a constraint system which meant that participants had to spend time grouping the pieces before they could rotate the half completed puzzle. The grouping of pieces required toggling a virtual button to enter the grouping mode, adding or removing the necessary pieces, and then toggling the virtual button again. The pieces could only be added or removed from the group via the Phantom Omni button and not with the glove. The clearance between pieces may have also contributed to longer times in the virtual training. Often, despite prior warning that the pieces did not have to be aligned perfectly, participants spent a lot of time trying to get the pieces in a near-perfect fit before moving on to the next piece. This led to excessive time spent on fitting single pieces as opposed to completing the puzzle as quickly as possible.

This was contrary to the PE where the pieces fit snugly and would stay in position. However, both hands were necessary for the assembly of the physical puzzle even during rotation as one hand was needed to hold the assembled puzzle. This was not required for the virtual training because gravity was not enabled.

We also observed that VE participants would favor one device over another during the familiarization session. For example, when participants struggled with selecting the pieces with the glove, they would then switch to the Phantom Omni and favor using this device. The same was true when they struggled initially with the Phantom Omni. This device preference may have influenced their performance and overall preferences.

In general we saw after the eight minute familiarization task, participants were sufficiently comfortable with the device use. Participants indicated that the VE was equally helpful as a training environment.

5.3 Role of Color

Training was done with colored pieces for both environments, but the PE group had the advantage of having handled the real puzzle. The PE group self-reported that their recall strategy was mainly shape and the VE group reported color as the main recall cue. Perhaps color is primarily used during the initial stages of learning when participants need to quickly identify which piece to be grabbing or examining. Selection of color is presumably a faster cognitive process than shape in this case. Since the PE participants reached the automaticity stage, it's possible that they relied more on shape than the VE group who may have relied more on color since they were still learning the assembly.

We believe VE participants spent most of their time in Fitt's cognitive stage learning the assembly of the puzzle and studying the color and shape of the pieces. For the PE group, they spent less time overall learning and more time in automaticity just putting the puzzle together. This difference in time spent in learning stages could be an explanation for the performance increase of the VE participants two weeks later. For the PE group, their learning may have been at a very short and shallow level while the VE group spent more time understanding the assembly process itself.

5.4 Individual Differences

We measured spatial ability of participants with a paper-based redrawn Vandenberg and Kuse Mental Rotation Task (MRT) [35]. Our study confirmed that there is a positive correlation between mental rotation ability and training performance. There was a significant difference in performance between the high and low MRT scoring participants. The fact that there were no significant differences between MRT scores and gender is contrary to findings of previous studies that males outperform females in MRT [36], especially since it was a paper-based MRT test.

Although significant differences were reported between genders for videogame experience, there was no evidence that videogame experience had a significant effect on task performance across groups. Technical expertise also had no significant effect on task performance across groups.

Participants reported that there was a significant difference in difficulty and ease of use with training environment across groups. However, there was no significant difference for self-reported realistic training environment and helpfulness of the training environment, which is promising when considering virtual reality as a training environment.

6 Limitations and Future Work

A number of limitations were identified during and after this study. These limitations and recommendations for future work are discussed briefly here.

One of the constraints and contributors to time taken in the VE group during training is the absence of a snap-to-fit function or constraint system [56]. The virtual puzzle pieces did not fit perfectly as there was some amount of tolerance between the pieces. Participants spent extra time attempting to fit the pieces perfectly despite being informed beforehand that the pieces will not fit perfectly. Additionally, once the piece was in the desired location, the participant had to take an extra step to group the pieces in order to rotate the semi-assembled puzzle to determine the fit for the next piece. A snap-to-fit function might also mimic to a certain degree the physical puzzles pieces which have a natural tight fit and stay in place after the pieces are fitted without having to first group the pieces. Of course, such a snap-to function would need to have a small boundary between pieces. The rationale in not providing a snap-to function was to mimic a real assembly environment as closely as possible. There is always a tradeoff in VR between utilizing technology and functionality that does not exist in the physical world which could possibly introduce extraneous variables, and attempting to keep it as similar as possible to the physical world.

The transparency of the selected puzzle pieces was another constraint that we faced. During the pilot phase the virtual representation of the glove was a solid color and when a puzzle piece intersected the glove, the piece turned slightly transparent indicating that it could be grabbed. However, since the solid colored virtual glove sometimes occluded the puzzle piece, we made the glove transparent for the study. Unfortunately, while the transparency of the virtual glove was a good solution, some participants complained that the transparency of the puzzle piece was not as discernable against the transparency of the glove. A recommendation for future studies would be to make selection of a puzzle piece in the virtual environment cause a complete change of color instead of just causing a change in transparency.

We observed individual differences for interaction between the glove and the haptic device. Participants showed a preference for either the glove or the haptic device for predominant use. The device preference seemed to depend on the equipment training phase where participants gained familiarity with the devices and interaction for 8 minutes. If participants experienced initial difficulty or comfort with either of the two devices, these preferences extended to the training and testing situations. We believe that in the future, participants should be encouraged and given tasks that use these devices equally and encourages bimanual operation. We speculate that this ambidexterity through bimanual usage will increase task performance speed.

We also observed that participants in the VE were pre-occupied with learning the devices and the virtual interface, learning the grouping procedure (subassembly) as well as concentrating on the fit of the puzzle piece. They did not seem to examine the instructions as often as the PE group. In testing, VE participants had difficulty remembering the assembly of the puzzle as was outlined in the instruction sheet.

In future studies, we recommend that another performance measure should be how often participants looked at the assembly instructions and at what time did they stop looking at the instructions. This could possibly lead to more understanding of the learning curve data and its relationship to Fitts' learning stages. This measurement could indicate what learning stage the VE participants spent the most time on. To complete the multi-modal approach, adding audio to match the sound of the wood blocks colliding would augment the haptic feedback and further immerse participants.

Another interesting factor to explore would be to determine if a participant's recall strategy changed from the initial test to the retention test. We would be particularly interested in what recall strategy the VE group used then since they reported using color as a cue more frequently than the PE group.

7 Conclusion

We have discussed a study conducted wherein we tested the differences in task and testing performance of a manual assembly of a six-piece burr puzzle between groups who were trained using physical objects and those who were trained with virtual objects.

We summarize our findings and contributions as follows. 1) While training in the virtual environment is outperformed by physical training on initial tests, appropriate use of color cues in virtual training can make it equally effective to physical training on performance testing after time has passed. 2) The learning curves within physical and virtual environments can differ, requiring different numbers of training iterations for each mode to avoid overlearning and premature skill degradation. 3) When designing training, it is important to consider not only the training time, but the anticipated time spent in Fitts' cognitive stage within that overall time.

Acknowledgments

This work was performed at the Virtual Reality Applications Center at Iowa State University as part of research supported by the National Science Foundation under grant # CMMI-0928774.

References

[1] C. Bailey, "Forgetting and the learning curve: A laboratory study," *Management Science*, vol. 35, no. 3, pp. 340–352, 1989.
[2] L. Yelle, "The learning curve: Historical review and comprehensive survey," *Decision Sciences*, vol. 10, no. 2, pp. 302–328, 1979.
[3] M. Oren, P. Carlson, S. Gilbert, and J. M. Vance, "Puzzle Assembly Training: Real World vs. Virtual Environment," in *Proceedings of the IEEE 2012 Virtual Reality Conference*. Orange County, California: IEEE, 2012, pp. 1–4.
[4] E. Thorndike and R. Woodworth, "The influence of improvement in one mental function upon the efficiency of other functions," *Psychological Review*, vol. 8, pp. 247–261, 1901.
[5] C. J. Hamblin, "Learning Transfer in Virtual Environments," *PhD Dissertation*, pp. 1–100, Jun. 2005.
[6] M. L. Gick and K. J. Holyoak, "The cognitive basis of knowledge transfer," in *Transfer of learning: Contemporary research and applications*, S. M. Cormier and J. D. Hagman, Eds. New York: Academic Press, 1987, pp. 9–46.
[7] J. A. Allen, R. T. Hays, and L. C. Buffardi, "Maintenance training simulator fidelity and individual differences in transfer of training," *Human Factors*, vol. 28, no. 5, pp. 497–509, 1986.
[8] J. Greenco, J.G.and Moore and D. Smith, "Transfer of situated learning." in *Transfer on trial: Intelligence, Cognition, and Instruction*, D. Dettermin and R. Sternberg, Eds. Norwood, NJ: Ablex Publishing Corp., 1993.
[9] R. J. Adams, D. Klowden, and B. Hannaford, "Virtual training for a manual assembly task," *Haptics-e*, pp. 1–7, 1999.
[10] P. Fitts, "Perceptual-motor skills learning," in *Categories of human learning*, A. Melton, Ed. New York: Academic Press, 1964.
[11] A. Boud, D. Haniff, C. Baber, and S. Steiner, "Virtual Reality and Augmented Reality as a Training Tool for Assembly Tasks," in *Proceedings of the IEE International Conference on Information Vizualization*, USA, 1999.
[12] R. Conway and A. Schultz, "The manufacturing progress function," *Journal of Industrial Engineering*, vol. 10, no. 1, pp. 39–53, 1959.
[13] J. R. Anderson, "Acquisition of cognitive skill," *Psychological Review*, vol. 89, no. 4, pp. 369–406, 1982.
[14] T. Wright, "Factors affecting the cost of airplanes," *Journal of Aeronautical Sciences*, vol. 12, no. 6, pp. 122–128, 1936.
[15] W. Arthur, W. Bennett, P. Stanush, and T. McNelly, "Factors that influence skill decay and retention: A quantitative review and analysis," *Human Performance*, vol. 11, no. 1, pp. 57–101, 1998.
[16] C. Lance, A. Parisi, W. Bennett, M. Teachout, D. Harville, and M. Welles, "Moderators of skill retention interval/performance decrement relationship in eight US Air Force enlisted specialties," *Human Performance*, vol. 11, no. 1, pp. 103–123, 1998.
[17] C. R. Hall, R. J. Stiles, and C. D. Horwitz, "Virtual Reality for Training: Evaluating Knowledge Retention," in *Virtual Reality Annual International Symposium*, Feb. 1998, pp. 1–6.
[18] K. Stanney, R. Mourant, and R. Kennedy, "Human factors issues in virtual environments: A review of the literature," *Presence*, vol. 7, no. 4, pp. 327–351, 1998.
[19] A. G. Gallagher, N. E. Seymour, J.-A. Jordan-Black, B. P. Bunting, K. McGlade, and R. M. Satava, "Prospective, Randomized Assessment of Transfer of Training (ToT) and Transfer Effectiveness Ratio (TER) of Virtual Reality Simulation Training for Laparoscopic Skill Acquisition," *Annals of Surgery*, vol. 257, no. 6, pp. 1025–1031, 2013.
[20] G. Rauter, R. Sigrist, C. Koch, F. Crivelli, M. van Raai, R. Riener, and P. Wolf, "Transfer of Complex Skill Learning from Virtual to Real Rowing," *PLoS ONE*, vol. 8, no. 12, pp. 1–18, 12 2013.
[21] J. Bliss, P. Tidwell, and M. Guest, "The effectiveness of virtual reality for administering spatial navigation training to firefighters," *Presence: Teleoperators and virtual environments*, vol. 6, pp. 73–86, 1997.
[22] C. Hall and C. Horwitz, "Virtual reality for training: evaluating retention of procedural knowledge," *The International Journal of Virtual Reality*, vol. 5, no. 1, 2001.
[23] J. Rodrguez, T. Gutirrez, E. J. Snchez, S. Casado, and I. Aguinaga, "Training of Procedural Tasks Through the Use of Virtual Reality and Direct Aids," in *Virtual Reality Environments*, C. S. Lanyi, Ed. InTech, 2012.
[24] S. Gerbaud, N. Mollet, F. Ganier, B. Arnaldi, and J. Tisseau, "Gvt: a platform to create virtual environments for procedural training," in *IEEE Virtual Reality Conference*, March 2008, pp. 225–232.
[25] B. Dalgarno and M. J. W. Lee, "What are the learning affordances of 3-d virtual environments?" *British Journal of Educational Technology*, vol. 41, no. 1, pp. 10–32, 2010.
[26] S. Moeser, "Cognitive mapping in a complex building," *Environment and Behavior*, vol. 20, pp. 21–49, 1988.
[27] B. Witmer, J. Bailey, and B. Knerr, "Virtual spaces and real world places: Transfer of route knowledge," *Int. J. Human–Computer Studies*, vol. 45, pp. 413–428, 1996.
[28] M. Piller and M. Sebrechts, "Spatial Learning in Transparent Virtual Environment," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 47, pp. 2133–2136, 2003.
[29] M. Prachyabrued and C. W. Borst, "Visual feedback for virtual grasping," in *IEEE Symposium on 3D User Interfaces*, Minneapolis, MN, 2014, pp. 1–8.
[30] D. Shuralyov and W. Stuerzlinger, "A 3d desktop puzzle assembly system," in *2011 IEEE Symposium on 3D User Interfaces (3DUI)*, march 2011, pp. 139–140.
[31] F. Ritter, T. Strothotte, O. Deussen, and B. Preim, "Virtual 3D Puzzles: A New Method for Exploring GeometricModelsinVR," *Projects in VR*, pp. 1–3, Aug. 2001.
[32] T. Parsons, P. Larson, K. Kratz, M. Thiebaux, B. Bluestein, G. Buckwalter, and A. Rizzo, "Sex differences in mental rotation and spatial rotation in a virtual environment," *Neuropsychologia*, vol. 42, pp. 555–562, 2004.
[33] P. Larson, A. Rizzo, J. Buckwalter, A. Van Rooyen, K. Kratz, U. Neumann, C. Kesselman, M. Thiebaux, and C. Van Der Zaag, "Gender issues in the use of virtual environments," *Cyberpsychology Behavior*, vol. 2, no. 2, pp. 113–23, 1999.
[34] A. A. Rizzo and J. G. Buckwalter, "Virtual reality and cognitive assessment and rehabilitation: the state of the art," *Virtual Reality in Neuro-Psycho-Physiology, Guiseppe Riva (Ed.)*, 1998.
[35] M. Peters, B. Laeng, K. Latham, M. Jackson, R. Zaiyouna, and C. Richardson, "A redrawn Vandenberg and Kuse mental rotations test: Different versions and factors that affect performance," *Brain and Cognition*, vol. 28, pp. 39–58, 1995.
[36] E. Waller and D. Knapp, "The transfer of spatial knowledge in virtual environment training," *Presence*, vol. 7, no. 2, pp. 129–143, 1998.
[37] K. Hinckley, R. Pausch, D. Proffitt, J. Patten, and N. Kassell, "Cooperative bimanual action," in *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1997, pp. 27–34.
[38] K. Hinckley, R. Pausch, D. Proffitt, and N. F. Kassell, "Two-handed virtual manipulation," *ACM Transactions on Computer-Human Interaction*, vol. 5, pp. 260–302, 1998.
[39] R. G. Marteniuk, C. L. MacKenzie, and D. M. Baba, "Bimanual movement control: Information processing and interaction effects," *The Quarterly Journal of Experimental Psychology*, vol. 36, no. 2, pp. 335–365, 1984.

[40] M. Peters, “Constraints in the performance of bimanual tasks and their expression in unskilled and skilled subjects,” *The Quarterly Journal of Experimental Psychology*, vol. 37, no. 2, pp. 171–196, 1985.
[41] R. A. Pavlik and J. M. Vance, “Expanding Haptic Workspace for Coupled-Object Manipulation,” in *ASME 2011 World Conference on Innovative Virtual Reality*. Milan, Italy: ASME, 2011.
[42] A. Bierbaum, C. Just, P. Hartling, K. Meinert, A. Baker, and C. Cruz-Neira, “Vr juggler: a virtual platform for virtual reality application development,” in *Virtual Reality, 2001. Proceedings. IEEE*, 2001, pp. 89–96.
[43] W. A. McNeely, K. D. Puterbaugh, and J. J. Troy, “Six degree-of-freedom haptic rendering using voxel sampling,” in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques - SIGGRAPH '99*. New York, New York, USA: ACM Press, 1999, pp. 401–408.
[44] R. A. Pavlik and J. M. Vance, “VR JuggLua: A Framework for VR Applications Combining Lua, OpenSceneGraph, and VR Juggler,” in *Workshop on Software Engineering and Architectures for Realtime Interactive Systems (SEARIS) in IEEE Virtual Reality*, Singapore, 2011.
[45] P. Khooshabeh and M. Hegarty, “Representations of Shape during Mental Rotation,” in *AAAI Spring Symposium: Cognitive Shape Processing*, 2010.
[46] G. W. Cobb, *Introduction to Design and Analysis of Experiments*. Springer, 1998.
[47] Ishihara, “Ishihara color blindness,” May 2013, http://www.toledo-bend.com/colorblind/Ishihara.asp.
[48] H. Ebbinghaus, *Memory: A Contribution to Experimental Psychology*. New York: Teachers College, 1913.
[49] K. J. Preacher, “Calculation for the test of the difference between two independent correlation coefficients [computer software],” May 2002, http://www.quantpsy.org/corrtest/corrtest.htm.
[50] J. Cohen and P. Cohen, *Applied multiple regression/correlation analysis for the behavioral sciences*, 2nd ed. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1983.
[51] S. R. Searle, F. M. Speed, and G. A. Milliken, “Population marginal means in the linear model: An alternative to least squares means,” *The American Statistician*, vol. 34, no. 4, pp. 216–221, 1980.
[52] J. M. Cortina and H. Nouri, *Effect Size for ANOVA Designs*. Sage Publications, 2000.
[53] J. Cohen, *Statistical Power Analysis for the Behavioral Science*. New Jersey: Lawrence Erlbaum Associates, 1988.
[54] F. Faul, E. Erdfelder, A. Buchner, and A.-G. Lang, “Statistical power analyses using g*power 3.1: Tests for correlation and regression analyses,” *Behavior Research Methods*, vol. 41, no. 4, pp. 1149–1160, 2009.
[55] P. Feltovich, R. Spiro, and R. Coulson, “Issues of expert flexibility in contexts characterized by complexity and change,” in *Expertise in Context: Human and Machine*, K. Ford and P. Feltovich, Eds. Cambridge, MA: MIT Press, 1993.
[56] J. H. O. Abhishek Seth, Judy M. Vance, “Combining dynamic modeling with geometric constraint management to support low clearance virtual manual assembly,” *Journal of Mechanical Design*, vol. 132, no. 8, pp. 1–7, 2010.

Patrick Carlson is currently a PhD candidate in Human-Computer Interaction (HCI) at Iowa State University. He received his Bachelors degree in Computer Science and Psychology at Simpson College. His research has focused on two areas. The first is evaluating virtual assembly training environments. The second is improving the socialization process of newcomers in open source communities. He is a member of ACM.

Anicia Peters is currently a Post Doctoral Scholar in the Department of Electrical Engineering and Computer Science at Oregon State University. She received her PhD and Masters in Human-Computer Interaction at Iowa State University, and her Bachelors degree from the Polytechnic of Namibia. She received a Fulbright Science and Technology award, a Schlumberger Faculty for the Future PhD scholarship and a Google Anita Borg Scholarship. Her research focused on digital signage and decision-making, and cross-cultural social media networks. She is a member of AIS, IEEE, and ACM.

Stephen B. Gilbert received a BSE from Princeton in 1992 and PhD from MIT in 1997. He is currently assistant professor in Industrial and Manufacturing Systems Engineering at Iowa State University, as well as Associate Director of the Virtual Reality Application Center and its Graduate Program in Human-Computer Interaction. His research focuses on technology to advance cognition, including interface design, intelligent tutoring systems, and cognitive engineering. He is a member of IEEE and ACM.

Judy M. Vance is the Joseph C. and Elizabeth A. Anderlik Professor of Engineering at Iowa State University and a Faculty Fellow of the Virtual Reality Applications Center. She received her B.S., M.S. and Ph.D. degrees in mechanical engineering from Iowa State University. Dr. Vance is past chair of the ASME Design Engineering Division and a former Associate Editor of the ASME Journal of Mechanical Design. Her research focuses on investigations of the use of virtual reality in product design and manufacturing. She is a member of IEEE, AIAA, ASEE and a Fellow of ASME.

Andy Luse is an assistant professor in Management Science and Information Systems at Oklahoma State University. Andy received his Ph.D. in Human-Computer Interaction and Computer Engineering from Iowa State University. Andy also received Masters degrees in Information Assurance, Computer Engineering, Business Administration, and Psychology from Iowa State University, as well as a Bachelors in Computer Science from Simpson College. His research interests include computer and network security, technology adoption, and statistical methods. He is a member of IEEE, ACM, and AIS.