ABSTRACT

MURPH, IMANI ALEXANDRIA. No Reaction to Distraction: An Investigation into Training Using Diminished Reality. (Under the direction of Dr. Anne Collins McLaughlin).

Combining learning theories and a novel technology, diminished reality (DR), may support training for complex tasks in distracting environments. DR can de-emphasize distractions in an environment and there may be a benefit in extending these effects to training methods. However, when a task is expected to be performed amid distractions, removing all distractions may not train the operator to ignore distraction during the actual task.

We hypothesized that using DR to gradually introduce the desirable difficulties of distraction will train the operator to perform in distracting environments, improve long term retention, and promote transfer of skills. 117 participants viewed two DR videos under three levels of distractions. The control group experienced all auditory and visual distractions, the smart DR group experienced reduced auditory and visual distraction, and the faded group started with a reduced amount of distractions, then ended with all distractions presence. Participants completed several measures of retention and transfer and individual differences in attention were also considered.

The use of DR videos, and specifically videos that scaffolding learners into learning under distraction was not an effective method for learning the ventilator assembly task, as measured by accuracy and completion time, the underlying concepts of the task, or for transferring that knowledge to a different context. Additionally, individual attentional differences were not impacted by type of DR experience. This study discusses previous research surrounding learning theories and considers the potential use of DR technology for effective training to achieve long term retention and transfer.

Keywords: distraction, learning, transfer, virtual reality, augmented reality

No Reaction to Distraction: An Investigation into Training Using Diminished Reality.

by

Imani Alexandria Murph

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Psychology

Raleigh, North Carolina
2024

APPROVED BY:

_____	_____
Dr. Anne Collins McLaughlin	Dr. Karen Chen
Chair of Advisory Committee	
_____	_____
Dr. Jing Feng	Dr. Shevaun D. Neupert

DEDICATION

To Walter and Sandra Murph.

BIOGRAPHY

Imani graduated from Florida Atlantic University in 2017 with a Bachelor of Arts in Psychology. She then completed a Masters of Science in Human Factors Psychology at Embry Riddle Aeronautical University. Immediately after, they continued their education at North Carolina State University. During her graduate school career, she has had multiple internship and research opportunities, including the Naval Postgraduate School, LeanGeeks, and most recently Microsoft, where she currently works as a User Researcher on the Xbox Research team.

ACKNOWLEDGMENTS

Thank you to my committee: Dr. Karen Chen, Dr. Shevaun D. Neupert, and Dr. Jing Feng for always providing a space to ask questions, to learn, and to grow. And a special thank you to my advisor, Dr. Anne Collins McLaughlin. Thank you for believing in my abilities as a researcher (and a rock climber) from day one. Your encouragement and your enthusiasm for research will continue to inspire me to be an even better researcher.

Thank you to my past and current labmates turned friends. You all have kept me sane through the roughest of semesters and the shortest of deadlines, even when those short deadlines were our fault.

Thank you to my family and to the friends that I've made on my graduate school journey. I'm grateful to have such a loving community to uplift me to achieve and do more.

Thank you to my Seattle Support System. In such a short amount of time, you have all shown me incredible support, care, and warmth in the past year.

Finally, thank you to my mom, Sandra Murph, for reminding me that I can always come home.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Whether learning complex flight systems or long division, humans face external interference on a daily basis. External interference can be defined as anything that might overload our working memory (Clapp et al, 2010). External interference can be divided into two categories: interruptions and distractions. Interruptions are important to completing the task and require the attention of the learner. For example, a loud, hard to ignore, low-speed warning in a cockpit requires a reaction from the operator. The operator needs to attend to the distraction, and then react to this interruption to continue safely flying the plane. Distractions, on the other hand, are irrelevant to the task and may derail performance, therefore, they need to be ignored, such as a phone going off during an exam or a monitor with extraneous information during a surgery. Tasks often have unavoidable interruptions and distractions (Clapp et al., 2010; Forster & Lavie, 2008). Over time, we learn to cope with it, a necessary part of learning, so that we are better equipped to deal with external interferences in a variety of situations.

Tasks, and learning how to accomplish them effectively, can also be broken down into two categories where performance can be measured: acquisition and retention. However, there is a phenomenon where performance at the time of learning (acquisition) and performance after a delay (retention) can differ; this is known as the paradox of retention. Retention of information can occur even in the absence of observable performance gains seen in acquisition (Soderstrom & Bjork, 2015). Learning (or retention) can be defined as a relatively permanent change in ability that results from practice or experience (Salmoni et al., 1984). In research studies of learning and practice, learners engage in an acquisition task, where an independent variable is manipulated. Common instructional manipulations applied during acquisition phases are the nature of instructions, the type of feedback, or the scheduling of practice (Schmidt & Bjork,

1992). A potential negative outcome of instructional manipulations, or additional aid during training, is an increased performance during acquisition but poor performance on retention tasks that take place after a delay. Essentially, having too much aid during the initial training period may give the appearance of learning, but the benefits of this may not persist after a period of time. This occurs when the instructional elements are highly supportive. However, when instructional elements are not as supportive during acquisition, the opposite outcome is displayed: lower performance during acquisition and higher performance on a retention task, leading to the paradox of retention (Schmidt & Bjork, 1992).

Once the target skill has been acquired, a further dimension of learning involves the ability to transfer what has been learned to new conditions and task variations (Seidler, 2010). While acquisition can show that a user has learned the target task well enough to perform without aid, transfer can demonstrate mastery of general concepts related to the target task and a broader understanding of the related concepts. Transfer of training, a type of learning, is usually desired in applied settings like education, where knowledge learned inside of a classroom could apply between school subjects and to tasks outside of scholastic settings (Adams, 1987). Transfer of training may also be desirable in dynamic environments such as healthcare or military applications, where a learner may have to adapt to new technology in an emergency, such as an unfamiliar cockpit layout.

Near and Far Transfer

Transfer can be further categorized into near transfer and far transfer (Kim & Lee, 2001). Near transfer emphasizes specific concepts and skills and occurs when there is a close match between the original training, target task content, and target task outcomes. Far transfer occurs when the original training, target task content, and target task outcomes do not align as closely.

An example of near transfer is a mechanic being able to work on cars of the same make, while far transfer would involve the mechanic's knowledge being broad enough to work on cars of different makes (Kim & Lee, 2001). A key 1986 study demonstrated far transfer when students enrolled in a college-level statistics course were able to apply their knowledge on statistical reasoning questions disguised as a household survey about sports. Students who were surveyed at the end of the course performed better than those surveyed at the start, showing that their knowledge from the statistics course transferred to a different context outside of the classroom (Fong et al., 1986). However, achieving far transfer is rare (Detterman & Sternberg, 1993) and requires the trainee to have a broader understanding of general concepts and skills (Kim & Lee, 2001). While acquisition can show that a user has learned the target task well enough to perform without aid, transfer can demonstrate mastery of general concepts related to the target task and a broader understanding of the related concepts.

Variability of Practice

Other studies have focused on ways to promote the same cognitive processes during training that will be required for a transfer test. For example, variability of practice can promote initial acquisition, which is necessary for achieving transfer (Schmidt & Bjork, 1992).

A recent study investigated variable practice on a laparoscopic surgery task (Vamadevan et al., 2024). Participants in the variable group practiced on four basic skills in a random order, and could not see what the next skill was. Those in the self-directed group were allowed to practice in any order. Researchers also prevented overtraining by removing a skill if participants reached proficiency, for both groups. An immediate transfer test was given, and 3-5 weeks later, participants took a retention test. Those in the variable practice group reached proficiency faster

during the basic skills phase, during the transfer test and during the retention test (Vamadevan et al., 2024)

By introducing variability in training, learners are better able to deal with variability in retention and transfer tests. Although, this may impair performance during practice, require more training time, and result in no improvement for previously practiced tasks, leading to the paradox of transfer. van Merriënboer et al. demonstrated this paradox by assigning them to a blocked practice or random practice group for a complex computer-simulated troubleshooting task (1997). The blocked group practiced four types of problems, grouped by type, 48 total problems practiced in blocks. The random group practiced all 48 problems in a random order, providing them with random practice. Both groups completed a four question test where two questions were previously seen (retention) and two questions were new problems (transfer). The results showed that the random group made more errors on the practice trials and had similar performance on retention test problems compared to the blocked group. However, on the transfer problems, the random group performed better than the blocked group (van Merriënboer et al., 1997). While the variability of problems in the random group allowed learners to extend the range of their learning to unfamiliar problems and develop a more flexible schema, it hindered their ability to perform well on the original task, as seen by the lower performance on the practice and retention questions. This is known as the paradox of transfer (van Merriënboer et al., 1997).

Practicing problems in a random order helped learners achieve transfer by introducing variability of practice. Variable practice forces the learner to use the same cognitive processes (by dealing with variability) during training that will be required by tests of retention or transfer. Despite initial performance being impeded, performance on a transfer task improves because the

same processes that were engaged during training are engaged during transfer. In sum, any manipulation that encourages the same forms of cognitive processing during training that will be required during retention or transfer can aid learning.

There are three specific theories of learning that can help explain the paradoxes of retention and transfer, and methods of instruction that stem from these learning theories.

Theories of Learning

Cognitive Load Theory

There are three types of cognitive load that learners can encounter: intrinsic, extraneous, and germane (Sweller et al., 1998). Intrinsic load is directly related to the task being learned, like driving a car, the simultaneous tasks of watching the road, steering the vehicle, and controlling speed. Extraneous load is when additional elements are involved in a task, but are not relevant to learning the task. Wickens et al. uses the example of having to look up abbreviations or dealing with technical issues as extraneous load when trying to learn a new software (2013). Germane load is a part of the learning process itself. For example, rehearsal of material, focusing on the meaning of material, or having to make choices (Wickens et al., 2013).

Distractions in a difficult and demanding task are considered extraneous load unless they are so much part of the task that performing with them is a necessary skill; even if they do not contribute to the outcome of the task. An example is when combat medics need to be able to filter out distractions, such as loud noises, even if doing so is not directly related to their medical duties. These distractions then become intrinsic load, as they are unavoidable. Workers should be able to work with these types of distractions. This is achieved with scaffolded training that exposes them to distraction as they are trained, but not so much that it impairs learning (Aoyagi

et al., 2019). This allows learners to get used to dealing with higher cognitive load, gradually, adding to their ability to handle or ignore unexpected distractions during a task.

Transfer Appropriate Processing

Transfer appropriate processing explains that explicit and implicit memory is facilitated by the degree of overlap between processes engaged during training and those engaged in retention tasks. These memory processes, which are critical for learning, are either conceptually driven (explicit) or data driven (implicit) (Franks et al., 2000). Although the original theory makes a distinction between conceptually driven and data driven processes, it is possible to create tasks where both types of processes are required (Blaxton, 1989). When it comes to complex tasks, learners are already using implicit and explicit memory. For example, when ignoring distractions they are using implicit memory and the actual task itself relies on explicit memory.

The benefits of variability of practice, as seen in van Merriënboer et al. (1997) and Vamadevan et al., (2024), can be extended to transfer appropriate processing. Learners in the random practice group have been exposed to variable training and engaged in the processes of retrieval necessary to succeed, thus they performed better on transfer tasks that involved variability. In regards to distractions, when learners are required to perform after experiencing distraction in acquisition, they are already engaged in the process of ignoring distractions along with performing the target task, making it easier to do both which also helps improve the explicit memory associated with the task (Blaxton, 1989). Nevertheless, too much distraction at the start of learning will make learning more difficult and overwhelm the learner. However, finding the right amount of distraction may be the key to help learners ignore them when they occur.

Desirable Difficulties

Another way to achieve learning and knowledge transfer is to introduce desirable difficulties during practice (Schmidt & Bjork, 1992). Difficulty is defined as any manipulation added to a task that degrades performance (Soderstrom & Bjork, 2015). While not all difficulties are desirable, when they trigger encoding and retrieval processes that support learning, they become desirable (Bjork & Bjork, 2020). Desirable difficulties were first applied in the literature on verbal learning and later extended to motor learning (Schmidt & Bjork, 1992).

One study investigated how desirable difficulties could improve recall of anatomy concepts (Dobson & Linderholm, 2015). In this study, desirable difficulties were created by adding a test condition which created a cognitively challenging environment, compared to the reading condition. Results showed that a read-test-read-test condition, where students had to complete a free recall task, was superior to conditions where students were required to passively read the concepts or complete a word fragment task between reading sessions. Students in the read-test-read-test condition recalled more information at the immediate and delayed (one week) assessment points (Dobson & Linderholm, 2015). Desirable difficulties differ from general difficulty in that they cause the learner to use active cognitive processing to overcome them when faced in acquisition, which leads to improved retention (Soderstrom & Bjork, 2015).

Cognitive load theory, transfer appropriate processing, and desirable difficulties describe basic aspects of learning, however, they do not inform us on their application, and used in isolation these theories may not be effective. For example, desirable difficulties that aren't intrinsic or germane load may add unnecessary difficulty without improving learning. Or, removing extraneous load during training that is normally present during the real task may make it harder to achieve transfer appropriate processing. There are, however, methods of instruction

that build upon these theories, combine the benefits of them, and demonstrate how to apply them for training purposes. The following section outlines these methods and how they connect back to these theories.

Methods of Instruction

Part Task vs Whole Task Training

Part task (PT) training is an instructional method where learners focus on subcomponents of a whole task (WT) to improve learning efficiency and overall retention on a whole task (Wightman & Lintern, 1985). For example, having a new pilot practice only controlling speed or only controlling altitude before attempting the whole task of piloting. Whole task training involves presenting a complete task to learners so they can practice the task as a single unit (Teague et al., 1994). One type of PT training is simplification, where a portion of the task is simplified by adjusting one or more of its characteristics. An example of this is keeping altitude steady during a flight simulator task, while the user learns to orient the plane. Tasks completed with the PT version, improves performance on the whole version of a task (Wightman & Lintern, 1985). However, PT training may not always introduce learners to the distractions that are a part of the WT, which leads learners to perform worse when there are distractions. This issue is further compounded by the fact that PT training separates tasks that are time-shared, or occur at the same time, in the whole task (Wickens et al., 2013). Thus, it does not allow learners to practice time-shared skills at the same time, which necessitates a further step of reintegrating subtasks into the whole task (Teague et al., 1994). Using the piloting example, a learner would not practice orienting and stabilizing the plane with PT training until they advance to the whole task.

An advantage of PT training is that it reduces the amount of intrinsic load experienced by learners, and allows for more cognitive resources to be dedicated to learning a specific aspect of the task (van Merriënboer & Sweller, 2005). Providing learners with an easier, less complex task can promote high performance initially, but lead to poor retention and transfer, as seen in the retention and transfer paradoxes (Schmidt & Bjork, 1992). Furthermore, PT training may be removing desirable difficulties by allowing learners to practice smaller portions of a more complex task. In regards to transfer appropriate processing, solely using PT training removes learners from the whole task context, which may make it harder to transfer knowledge when given a retention test on the whole task. The aforementioned theories of learning and their influence on learning demonstrate that learners need a balance between PT and WT training, rather than one versus the other.

Emphasis Change as a balance between PT and WT

Emphasis Change Protocol (ECP) is a theory of instruction similar to PT training. However, with ECP, the learner's attention is directed to a specific portion of the task, while still experiencing the whole task (Gopher et al., 2007). For example, a learner completes the task of driving as a whole, but is told that their priority is watching for pedestrians, rather than controlling speed. ECP is also known as variable priority training which has been shown to accelerate learning and increase task mastery (Boot et al., 2010).

Kramer et. al. (1995) investigated the effectiveness of variable priority training compared to equal priority training for younger and older adults on a dual task paradigm involving monitoring task and an alphabet arithmetic test. All participants were able to see their in-the-moment performance via a feedback graph. Participants in the variable priority condition were told to focus their efforts for the monitoring task and the alphabet arithmetic tasks with the

following percentages: 80-20, 35-65, 50-50, 65-35, 20-80. These percentages varied between blocks and were communicated to participants via a vertical line on the feedback graph. Those in the equal priority training condition focused their efforts 50-50 for both tasks. In this study, emphasis change was accomplished via instructions (i.e. telling the participant to focus on one aspect of the task) and via performance indicators for the subtask (i.e. the feedback graph being emphasized) (Kramer et al., 1995; Gopher, 2007). Results showed that variable priority training is effective for memory based tasks, in addition to the psychomotor tasks that have been used in past research. Variable priority training also led to automatization of the two tasks as well as a generalizable skill of dual task management. This generalizable skill transferred to a novel dual task paradigm. The benefit of emphasis change, when compared to PT training, is that rather than the task being broken into smaller parts, learners still experience the whole task. This avoids the issue of breaking down a task in conceptually meaningful ways, and then re-integrating the steps (Gopher, 2007).

Additionally, under the theory of transfer appropriate processing, it is more beneficial for learners to maintain the whole task context, distractions included, so when tested on the target task, they have already experienced the processes needed to ignore those distractions.

Scaffolding

Another method of improving learning is scaffolding, which refers to the gradual removal of aid or instructions during training while increasing the difficulty of the task. This is because instructional aids become redundant as learners gain more knowledge (Herbert et al., 2021). Emphasis change is a way to gradually increase the amount of desirable difficulty, or distractions, that a learner faces by having distractions become more apparent while learners are still required to complete the whole task. Additionally, emphasis change can scaffold learners to

the intrinsic load of distraction, rather than exposing them to it all at once. Augmented reality (AR) technology can be leveraged to create training programs based on these theories of learning and instruction. An example can be seen in a study by Hughes et al. 2022, where 10 surgeons and 10 medical students completed a simulated laparoscopic surgery task under a control condition with no distractions, and then completed the same experimental condition with auditory and visual distractions, which essentially scaffolded learners to being able to work under distraction. The medical students had little experience in laparoscopic surgery. In the control condition, surgeons performed better than students. However, in the distraction conditions, on average, surgeons' completion time was higher compared to the control condition. Students' completion time, on average, stayed the same and in some cases decreased, possibly due to being novices at the task (Hughes et al., 2022). Additionally, Herbert et al. used AR and emphasis change to gradually increase desirable difficulties by removing instructions and redundant information from the AR system which lowered the amount of cognitive load experienced by learners. When learning to manage distraction in task performance, a novel form of AR, called Diminished Reality (DR), may be helpful in controlling exposure to distraction.

Diminished Reality

DR is a form of augmented reality (AR); AR is a technology where interactive digital objects are overlaid into the user's environment (Azuma, 1997). A more recent definition defines AR as "a technology that extends the user's reality using digital information" (Eckert et al., 2019). For example, AR may help with medical visualization, where practitioners can visualize, manipulate, and overlay patient images to create an interactive, immersive environment (Murali et al. 2021). There are also training applications, where a learner may see instructions on how to complete a task overlaid into their environment without needing to look away from the task

(Azuma, 1997). AR can also involve adding auditory information into the environment, such as an audio tour in a museum that provides information based on what the user is looking at (Yao & Liang, 2023). AR has also been used to reduce students' cognitive load in science reading by utilizing the contiguity effect, where pictures and related text are displayed simultaneously (Lai et al., 2019).

Rather than adding to the environment, DR removes visual and auditory elements. DR can, more specifically, remove distracting elements of the environment. This is similar to the concept of decluttering, where elements that are of less importance to the main task are made less salient compared to surrounding elements via methods such as dimming, size reduction, turning symbols into dots (dotting), or complete removal from visual displays (St. John et al., 2005). Decluttering via dimming and dotting on static displays minimized visual complexity, resulting in improved visual search and target performance for both novice and expert users (Lazaro et al., 2021). Decluttering has been limited to elements on electronic displays, such as monitors (O'Hara et al., 2007). DR can be applied to the physical world and digitally remove or obscure large objects or even people, making it easier to focus on the current task.

DR's unique ability to gradually remove elements from the environment and direct attention may serve as a method to apply multiple theories of learning into one training program to promote long-term retention and transfer of skills to a new context.

The Design of DR

When considering what elements of an environment are distracting, it is helpful to understand how preattentive features guide attention (Wolfe, 2014). Knowing which basic elements of a scene capture attention helps in the creation of distractions that are salient, which refers to the conspicuity or contrast of an item relative to its surrounding environment (Itti &

Koch, 2001). This is particularly useful when a DR designer has to create visual distractions, for training purposes, or reduce existing distractions within a virtual environment to allow a learner to focus on the main task. Wolfe (2014) discussed the general rules that dictate if a feature will capture attention during visual search. While Wolfe focused on the visual sense, some of these rules can be applied to other sensory modalities. One of the rules states that the greater the difference between the target and the distractors, the more efficient the search, and the more likely that the target will capture attention (Wolfe, 2014). For example, in a video game, quickly aiming a reticle at the correct target would be easier if the target stood out visually from the environment, via color, movement or other visual qualities. When it comes to DR, to make a distracting object less salient, a designer would need to create a difference between the distracting objects and the task related objects. Yao et al. (2013) accomplished this by blurring background elements in a video conference call making them less salient compared to the current speaker who was in full focus. This has also been accomplished via decluttering by dimming airplane symbols during a threat detection task, making these symbols harder to attend to (St. John et al., 2005).

Designing DR using principles of human attention is necessary to capture attention for important task elements and create effective diminishment of non-important task elements. Both are required to create training that utilizes emphasis change.

The blanket application of DR involves removing all distractions from a task. This method is similar to part task training, by removing the distracting portion of the task and therefore reducing the learners' cognitive load. However, when distractions are a part of a task, according to transfer appropriate processing, learners need exposure to them during knowledge acquisition to benefit retention. Moreover, removal of distractions removes the desirable

difficulty associated with tuning them out, while too many distractions can frustrate learners during training. When collectively applying these theories to DR and gradually placing more emphasis on ignoring distractions, it may be possible to deliver effective training to achieve retention and transfer on complex tasks.

Distance Learning, Simulators, and Distractions

To understand whether real life simulation training with distractions improves students' ability to handle distractions Thomas et. al. (2015) conducted a study in which medical students completed a simulated ward round twice, with a 4 week interval in between. In the first ward round, distractions (task irrelevant) and four interruptions (task relevant) were present. Targeted, individualized feedback on how to handle the distractions was given to the experimental group immediately after the first round. The second round served as the post intervention. Both groups had reduced error rates during the post intervention round, and the group that received feedback had the lowest error rate. Results from the study that learning the skills to manage distractions is essential (Thomas et al., 2015). Medical students may study under varying levels of distraction, but typically, these locations may not involve the same amount or intensity of distractions. A medical student studying for clinicals may study in a library, or a busy coffee shop, but the operating room may bring about unexpected and more intense distractions. Even final year medical students are not "inherently equipped" with the skills needed to handle distraction and avoid errors in a medical environment (Thomas et al., 2015).

Distractions are inherently incorporated into real-time learning environments, sometimes incorporated into simulators, but there are benefits to incorporating them into distance learning modules, as well, which would allow students to learn to work with these distractions, in addition to the benefits associated with distance learning.

In another study, researchers found no difference in learning outcomes when comparing online vs in-person lectures; however, they found that it negatively impacts engagement with classmates and lecturers, which contributes to the overall satisfaction of the learning experience (Sandrone et al., 2021). Researchers in this study made a distinction between passive and active learning, where passive involves sitting in a classroom and listening to a lecture, and active learning involves "doing things and thinking about the things they are doing" (Bonwell & Eison, 1991). This same study mentions ways that active learning can be incorporated into online, asynchronous learning, by using tools like Panopto to insert quizzes into video lectures. Asynchronous learning allows for both passive and active learning to take place, which allows students to engage in active learning during live lectures, with instructors adjusting their teaching based on students' understanding of the concepts being taught (Sandrone et al., 2021). Morgan et al. (2002) compared learning outcomes between students who had experiential learning via a medical mannequin or visual learning via an educational video. Comparisons of pre and post test scores showed a significant improvement with no significant difference between video and experiential learning groups in those scores, or in the final examination scores. In a more recent study, researchers investigated whether combining an online video assisted teaching program with traditional classroom learning leads to better test scores compared to students who only partook in traditional classroom learning. Results showed that the online video assisted groups showed better scores in post test scores and in mean OSCE (objective structured clinical examination) scores compared to the traditional classroom group. Additionally, students in the video assisted group self reported feeling more confident in the procedure (Chucherd et al., 2023). In both Chuchard et al. 2023 and Morgan et al. 2002, researchers came to the conclusion that video assisted learning is a useful tool to aid traditional learning methods.

Is VR Enough?

While VR and AR can be powerful tools for training, they are not without limitations. Across two different literature reviews of 14 medical simulators, these studies note that simulators do not lead to transfer of skill to real life and that they need to be supplemented with evidence based learning methods (Lee et al., 2020; Papanikolaou et al., 2019). Another study discusses higher fidelity systems and how they benefit experts, but lower fidelity systems are more likely to benefit novice learners. In general, more research is needed on the connection between expertise level and the benefits of high or low simulator fidelity (Lefor et al., 2020). Novice learners need a place to start, and while the difficulty of a high fidelity simulator may be beneficial for experts, it makes learning harder for novice learners (Bjork & Bjork, 2020). More expensive, higher fidelity simulators may not be necessary to achieve desired learning outcomes, and may negatively impact these outcomes. Another limitation is the lack of ability to train remotely on VR or AR tools. Users learning from home, or other locations, can't access complex and expensive VR/AR equipment.

As VR becomes a more common learning tool, there is more research on how to improve these systems. However, we should not turn our focus away from 3D video instruction and remote methods of learning in general, as these areas can still benefit from research to incorporate distractions that mimic real world environments to train learners, even remotely, on how to work with distractions and therefore prepare them for real-world environments.

Study Overview

The current study aimed to investigate how creating emphasis change and desirable difficulties via DR in training videos, may influence retention of a complex medical task and how well the retained skills transfer to a different context. Generally, we expected performance

to increase when gradually introduced to distractions. We measured performance via a retention and transfer test, as well as concept tests based on acquisition videos. We also examined how individual differences may impact the effectiveness of faded DR.

Types of DR

Participants experienced one of two types of DR (smart or faded) or a control condition. In the control condition, auditory and visual distractions were fully present, such as loud alarms, irrelevant conversations, as well as people and objects floating past during the video. An image of what the control condition looks like can be seen in Figure 1. The task steps said aloud in the video were delivered via a voice over of someone stumbling over their words and backtracking on incorrect answers so that the delivery of the steps themselves was as distracting as possible. Additionally, the camera view would turn to look at the distracting objects floating by during the video. With smart DR, participants experienced blurred visual distractions, with reduced auditory distractions. A screenshot of what the smart DR condition looked like can be seen in Figure 2. The task steps being spoken aloud in the smart DR condition were spoken clearly and deliberately. With the second type, faded DR, participants watched the assembly task with smart DR for the first acquisition video and watched the control condition for the second acquisition video.

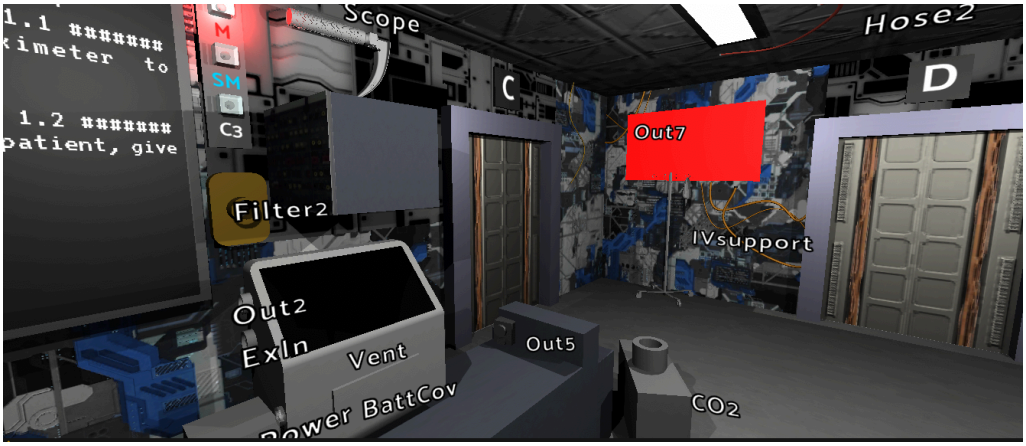


Figure 1. A screenshot of the control condition in the acquisition videos.

Figure 2. A screenshot of the smart DR condition in the acquisition videos.

Pilot Study

Method

Participants

A remote pilot study was conducted to assess the difficulty of the experimental tasks. A total of 15 participants completed the pilot study, and were recruited from NCSU Cognitive Psychology courses. Participants were either compensated with course credit or $20 upon

completing both days of the study. We screened for motion sickness, epilepsy, visual impairment, and hearing impairment. Participants were also disqualified from the study if they participated in research involving Diminished Reality in the past 6 months.

Materials

Informed Consent. Participants consented to the study via an IRB-approved form that explained study details, eligibility requirements and compensation.

Demographics. Participants filled out information about their age, gender, familiarity with AR/VR technology, and their medical knowledge. The medical knowledge survey has been used in Hicks et al. (2018) and Murph et al. (2021) to determine participants' prior knowledge of ventilators and their uses. In the pilot, we only looked at knowledge of medical ventilator responses.

Desktop VR Application. The Desktop VR application was used by the experimenter and the participant. The term Desktop VR in the current pilot refers to viewing a computer generated VR environment in full screen that can also be used in a headset, but that is being experienced on a desktop, with the ability to look around the environment freely. The experimenter, using their keyboard, was able to control the progression of the task, inform participants of incorrect answers, control which experimental condition would be experienced and pause the task if needed. When participants gave an incorrect answer, they heard a sound indicating that the answer was incorrect and were told to try again. The participant was able to look around the environment using their mouse.

Zoom Video Conferencing. Zoom Video Conferencing was used to communicate with the participants as they completed the task. Additionally, this allowed the experimenter to hear participants' answers and progress the task.

NASA Task Load Index (TLX). The NASA Task Load Index was initially created in 1988, with six subscales: mental demand, physical demand, temporal demand, performance, effort and frustration (Hart & Staveland, 1988). The subscales accounted for a significant amount of variance in the sample, with an overall r^2 value of .86 ($p < .01$).

The NASA TLX has also been evaluated for reliability in a healthcare setting ($\alpha = .72$), with a sample of 757 ICU Nurses (Hoonakker et al., 2011). This study also found that the TLX had high convergent validity and correlated highly with the Perceived Workload Scale ($r = .77$) and the Rating Scale Mental Effort ($r = .81$).

Design

Levels of DR were manipulated between participants, where each participant experienced one of three types of DR. Level of DR was the independent variable, while accuracy on the two acquisition tasks and retention task were the dependent variables, in addition to completion time. Measures of workload were also collected.

Experimental Tasks

Acquisition 1 and 2. On the first day of the experiment, participants, using "voice commands" assembled a medical ventilator, twice. As participants said the correct steps, the experimenter progressed the task. During the acquisition phase, participants experienced one of the three types of DR. They had the ability, using their mouse, to look around the environment and find the necessary parts. There was an instruction screen within the environment that listed out the correct steps, however, the parts needed were labeled differently. For example, the pulse oximeter was labeled as "oxypulse" and the A/C outlet was labeled "ElectricOut." An example of how the different conditions looked in both acquisition phases can be seen in Figure 3.

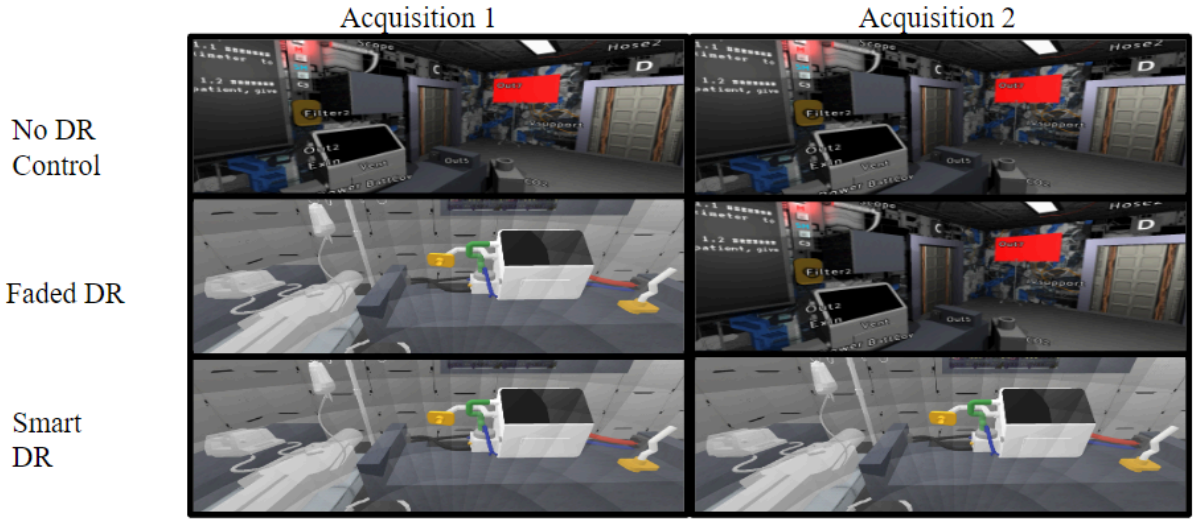


Figure 3. An image of the three types of DR, across acquisition 1 and 2 in the pilot study.

Retention Task. One week later, participants video conferenced with the experimenter and had to complete the same assembly task. However, the instruction screen was blank, and participants had to assemble the parts from memory. During the task, all participants experienced the control condition, with all visual and auditory distractions present.

Procedure

Prior to starting the experiment, participants completed demographic information and downloaded the Desktop VR Information. Then, participants logged onto a zoom call with the experimenter, were randomly assigned into one of the three conditions, and completed acquisition 1 and 2, and completed the NASA TLX after each trial. One week later, participants logged onto a zoom call again, and completed the retention task, followed by a final workload assessment. After, participants were either compensated via course credit or with $20 via their preferred payment app.

Pilot Study Results

All participants were between 18 and 24 years old. Eight women and seven men participated. Only one participant reported having "a lot" of experience with VR, while seven reported "some" and the remaining seven reported having no VR experience. Finally, for knowledge of medical ventilation procedures, zero participants had good or excellent knowledge, while four had average knowledge, five had poor knowledge and six had no knowledge.

Inferential statistical analysis was not conducted due to the sample size and pilot nature of the study. Descriptive statistics of the dependent variables can be found in Table 1. During the final retention trial, only one participant was able to complete the task in the allotted time. Completion time by group, across trials can be seen in Figure 4.

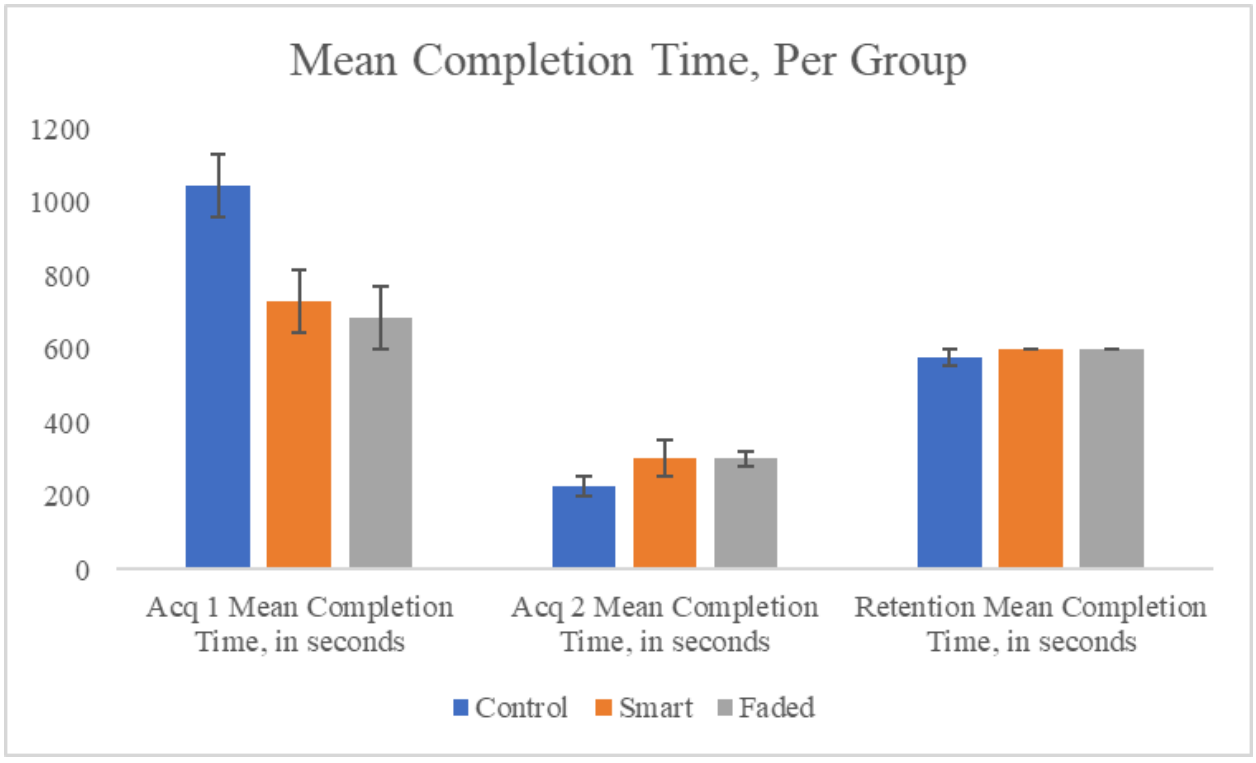


Figure 4. Mean completion time in seconds, per group, across trials Error bars represent standard errors of the mean (SEM).

Large standard deviations in errors were due to differences in guess-rates; some participants gave rapid fire guesses when unsure, while others gave slower, more deliberate guesses. Amount of errors by condition across trials can be seen in Figure 5.

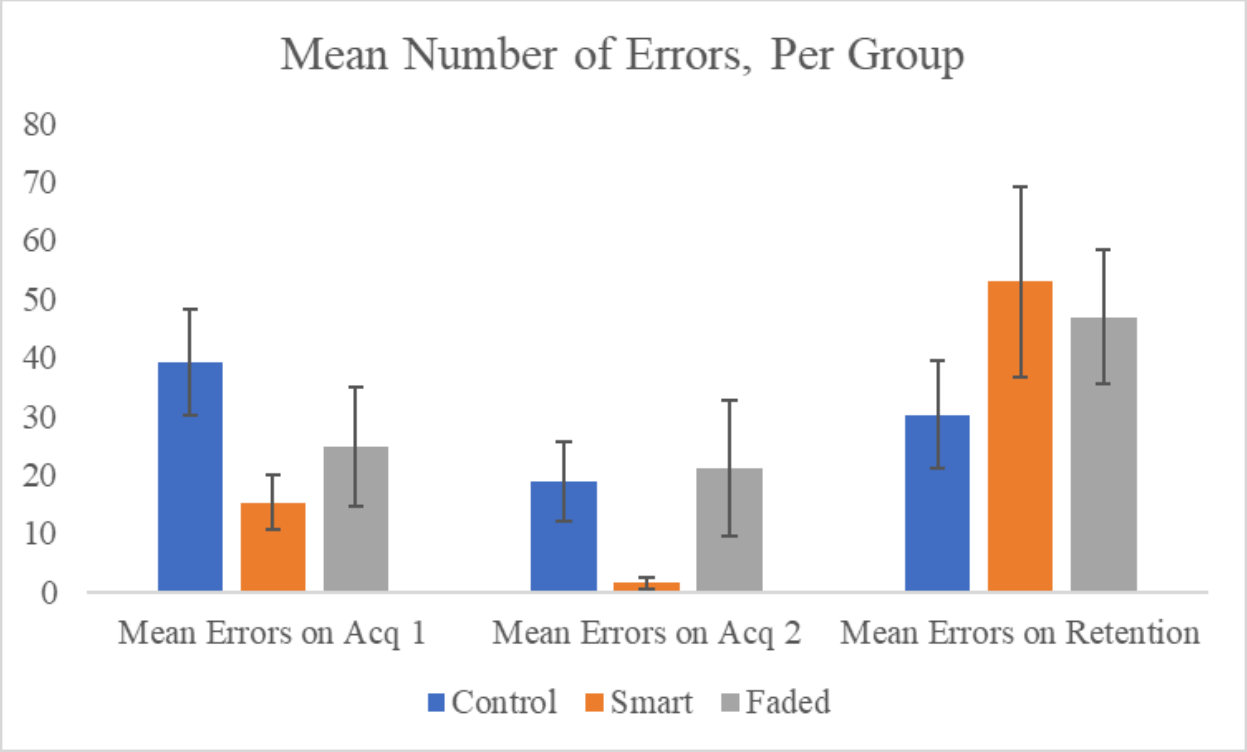


Figure 5. Mean number of errors, per group, across trials. Error bars represent standard errors of the mean (SEM).

Finally, workload did change across conditions, with lower average workload ratings occurring in acquisition 2. Average workload ratings across trials can be seen in Figure 6.

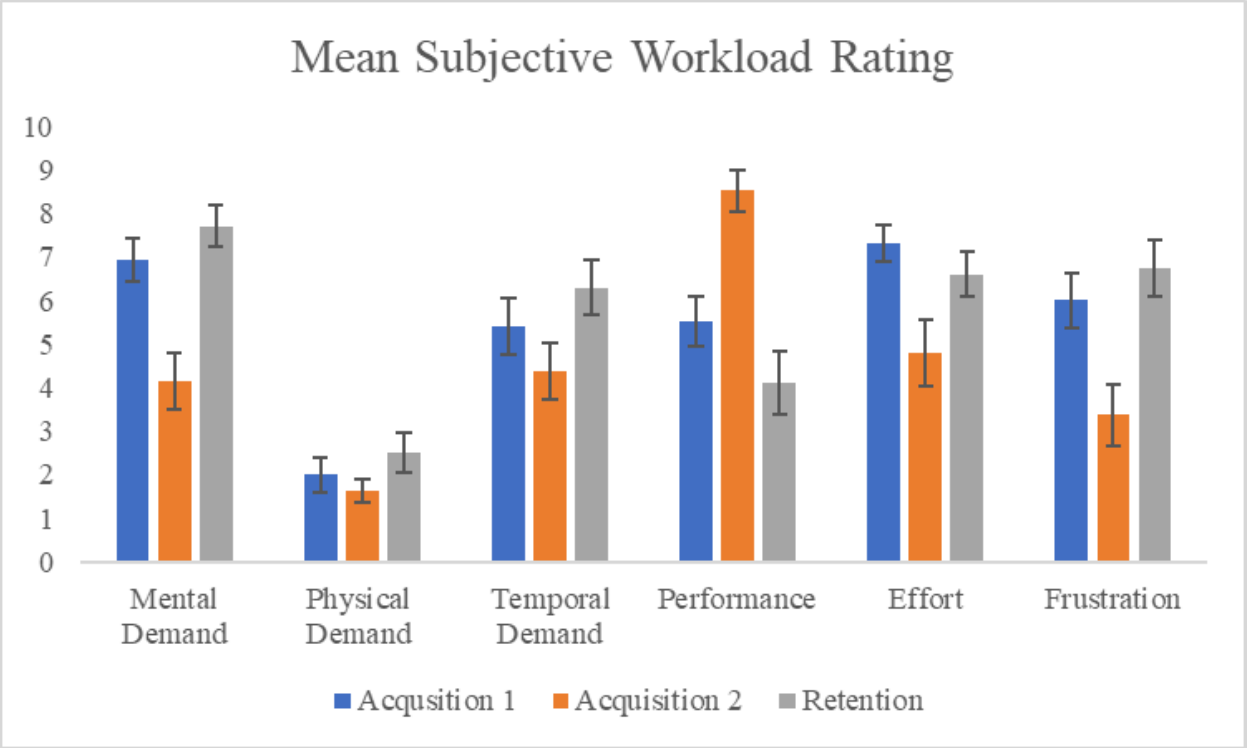


Figure 6. Average workload ratings, across trials. Error bars represent standard errors of the mean (SEM).

Table 1.

Pilot Study Results

	n=7 Control		*n*=4 smart DR		*n*=4 Faded DR		*n*=15 All Participants	
	M	SD	M	SD	M	SD	M	SD
# of counted Errors in Acq 1	39.14	22.13	15.25	8.10	24.75	17.46	28.93	20.75
# of counted Errors in Acq 2	18.86	16.55	1.50	1.66	21.25	20.01	14.87	17.36
# of counted Errors in Retention	30.29	22.81	53.00	28.09	47.00	19.69	40.80	25.65

Discussion

In general, participants did poorly on both acquisition trials and the retention trial and did not learn the assembly task. The mean number of errors did decrease in acquisition 2, but increased during the retention trial. Completion time somewhat followed this trend, however, participants were limited to ten minutes in the retention trial. All but one participant managed to complete the retention task in the time allotted. Finally, across measures of workload, sub-scales such as mental demand and frustration were high during acquisition 1, decreased during acquisition 2, but increased again during retention. The opposite effect was seen on the performance subscale.

Overall, the interactive assembly task was difficult, especially during retention across all conditions. Participants may have had difficulty due to having to recall the steps from memory, without any instructions provided. Additionally, error rates varied based on how participants guessed the correct steps. Some participants used a rapid-fire guessing method, while others used a slower, more deliberate guessing method.

Additionally, with instructional aid, such as the task steps being present, may have explained the difference in performance between acquisition 2 and the retention test, especially considering the one week interval. The increase in errors on retention may be because there was too much aid in acquisition 2, and the benefits of the additional aid (having the task steps present), likely did not persist after a period of time. (Schmidt & Bjork, 1992).

Between the pilot test and the main study, three main changes were made. First, the acquisition task was changed to a passive retention test, rather than an active task. Participants were also not given feedback on whether their answers were correct. We also reduced the

interval between acquisition and retention. Finally, the retention task was changed from an active task to a passive retention test.

Study 1 Hypotheses

Predictions Regarding Performance on Retention and Transfer Tests

H1A: There will be differences in the effects of DR type on retention test accuracy. Contrasts will reveal significant differences in accuracy and completion time on a retention test with significant differences between smart DR and faded DR, as well as differences between the control and faded DR, with faded DR leading to the highest accuracy in both comparisons.

H1B: There will be differences in the effects of DR type on transfer test accuracy. Contrasts will reveal significant differences in accuracy and completion time on a transfer test with significant differences between smart DR and faded DR, as well as differences between the control and faded DR, with faded DR leading to the highest accuracy in both comparisons.

HIC: There will be differences in the effects of DR type on concept test accuracy. Contrasts will reveal significant differences in accuracy and completion time for concept tests with significant differences between smart DR and faded DR, as well as differences between the control and faded DR, with faded DR leading to the highest accuracy in both comparisons.

Predictions Regarding the Effects of Individual Differences on Retention and Transfer

H3A: There will be a strong positive relationship between high Adult ADHD scores and lower performance, as measured by accuracy and completion time on the second

retention test, while also showing that those with high ADHD scores receive more of a performance benefit when in the faded-DR condition compared to the control.

H3B: There will be a strong positive relationship between Attentional Control Scale Scores and better performance, as measured by accuracy and completion time on the second retention test, while also showing that those with lower attentional scores receive more of a performance benefit when in the faded-DR condition compared to the control.

Method

Participants

A power analysis was conducted to determine sample size using G*Power (Faul et al., 2009). 117 participants were needed to achieve an .80 power level, with a medium effect size (Cohen's d = .5) and a Type I error probability of .05. A total of 172 participants qualified for the study. Recruitment occurred on a rolling basis via North Carolina State University's SONA system, which provides course credit as compensation to Psychology 100 students for their participation in research studies. We also screened for motion sickness, epilepsy, visual impairment, and hearing impairment. Participants were disqualified from the study if they participated in research involving Diminished Reality in the past 6 months. They were also required to use Safari or Firefox in private browsing mode, which was necessary for the embedded distractor audio to play during the retention and transfer tests. Further demographic information can be found in Table 2. After each retention test and the transfer test, participants were asked if they heard audio. If they reported hearing audio in the background of the retention and transfer tests, their results were included in the study, leaving 136 participants. The other 36 participants either heard no audio, or only audio on some of the retention or transfer tests, which may have been caused by autoplay issues with their browser. Finally, as a quality control

measure, outliers in total completion time were removed from the sample if they completed the first retention or transfer test in less than 30 seconds.

Materials

Informed Consent

Participants consented to the study via an IRB-approved consent form that outlined study details, eligibility requirements, and compensation.

Demographics

Participants filled out information about their age, gender, English proficiency, familiarity with AR/VR technology, and their medical knowledge prior to the experiment. The medical knowledge survey was the same survey that was used in the pilot study.

Adult ADHD Self Report Scale (ASRS)

The 18 item Adult ADHD scale was used as a measure of subjective attentional control, with a score range of 18 to 90. This scale was placed between the two retention trials to serve as a filled interval to counteract any short-term memory benefits associated with recent experience of the task. It also provided general self-report information about the attentional abilities of the sample and whether ability levels were randomly distributed between the manipulated conditions, as a manipulation check.

The scale was initially validated in 2005 with 154 participants (Kessler et al., 2005). The 18 item scale was found to have a total classification accuracy, defined as the ratio of the number of correct predictions to the total number of input samples, of 96.2%. The scale was also validated more recently, with 646 clinically diagnosed adult ADHD patients and 908 population controls (Brevik et al., 2020). Internal consistency measured by Cronbach's alpha was 0.952

for the full ASRS scale. Additionally, a principal component analysis was conducted, and two factors, inattention and hyperactivity, explained 62.2% of the variance in the sample.

Attentional Control Scale (ACS)

The Attentional Control Scale is a 20 item self report scale, with a score range of 20 to 80, and was used as an additional measure of attentional control (Derryberry & Reed, 2002). Participants took the scale after completing demographic information. It also provided general self-report information about the attentional abilities of the sample and whether ability levels were randomly distributed between the manipulated conditions.

In 2014, the scale was further validated with a confirmatory factor analysis, with the two factors being attentional focusing and attentional shifting, explaining 33% of the variance. Cronbach's alpha for focusing was reported as .82, and shifting was reported with a Cronbach's alpha of .71 (Judah et al., 2014).

Design

Levels of DR were manipulated between participants, where each participant experienced one of three types of DR. Level of DR was the independent variable, while accuracy on retention tests, transfer tests and concept tests were the dependent variables, in addition to completion time.

To assess differences in mean accuracy and completion time between groups and test hypotheses H1A-H1C, we will run two sample t-tests, with an alpha level of $p > .05$, between faded DR and control, as well as smart DR and control. Based on the hypotheses, we are specifically interested in assessing performance benefits of faded DR compared to the two other groups. For hypotheses H2A-H2B, we will run multivariate linear regression analyses to assess the relationship between attentional control, the type of DR experienced and performance.

Experimental Tasks

Acquisition Videos of VR Assembly Task

Three videos for each condition were used where participants were expected to learn the steps to the ventilator assembly. Figure 7 shows what the ventilator looks like. Each individual video matched one of the levels of DR, and showed the assembly task occurring, with audio overlaid with a voice saying the correct ventilator steps. The VR environment videos were based upon the environment used in Murph et al. (2021). The VR environment videos also included visual and auditory distractions based upon which condition participants are in. The VR ventilator assembly task was adapted from the ventilator simulation task used in Hicks et al. 2018 that used real medical tubing and props. There are 23 task steps, which can be seen in Appendix A.

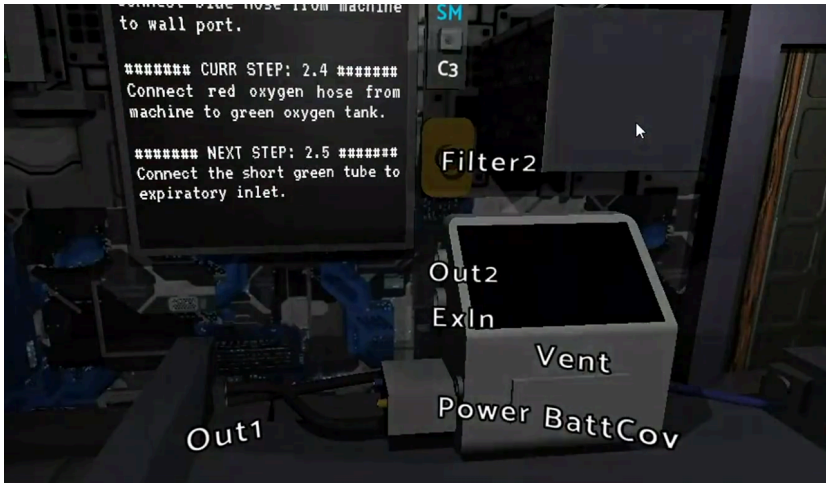


Figure 7. A screenshot of the ventilator used in the acquisition videos.

Retention Ranking Test

The ranking test was based off of the steps said aloud in the acquisition videos. Participants were required to put the task steps in order based on their memory of the videos. A list of task steps can be found in Appendix B. No feedback was given.

The retention test had a 6 minute and 30 second time limit. The time limit was based on the length of a distracting audio clip pulled from McLaughlin et al. (in preparation), that was related to the VR environment shown in the videos, and included similar distractions as the control condition. In addition to correct rank order, we also scored participants based on distance from the correct answer and the longest correct sequence. These two measures were calculated after data collection due to seeing poor performance on the pure ranking scoring method. An explanation of these scoring measures can be found in Appendix K. Last, completion time was collected to determine how long participants spent ordering the steps.

Transfer Ranking Test

In the transfer ranking test, participants saw a list of steps and put them in order based on the order they learned in the acquisition videos. A list of task steps can be found in Appendix C. Participants were required to remember the steps from the acquisition videos, as no instructions were given during the transfer ranking test. The steps were created to be similar to the acquisition trial and used the same syntax, but used different wording. All participants experienced the same distracting audio that was experienced during the rank order retention test and had the same 6 minute and 30 second time limit. The test was scored similarly to the Retention Ranking Test and additional information on the distance and longest sequence scoring methods can be found in Appendix K.

Multiple Choice Concept Tests (Retention and Transfer)

Participants started with a pre-test that required conceptual knowledge of the ventilator assembly shown in the videos. Because the ventilator was not based on a real one, but a simulated ventilator from Hicks et al. 2018, it was expected that prior knowledge of medical ventilators would not provide any additional assistance during the first concept test. The second

concept test was based on the rank order transfer test and had the same syntax, but different wording. Third concept test was the same as the first, with questions presented in a randomized order. Concept tests can be seen in Appendix D and E. Concept tests were created based off of audio said aloud in the VR videos, information that flashed on the ventilator screen and required participants to understand the reasoning behind the order of the steps.

Procedure

Study responses were collected via Qualtrics and distributed via the SONA recruitment pool at North Carolina State University. Prior to completing the study, participants provided informed consent via Qualtrics, completed demographic information, the ACS, and then took the first ventilator concept test. Based on their assigned condition, participants then watched the control, smart DR, or faded DR videos. They were not allowed to continue until the video had finished playing. After watching the videos, participants took the first rank order retention test. Once the test was completed, or the time limit expired, participants moved onto the rank order transfer test. Following the rank order transfer test, participants completed the transfer concept test, the medical knowledge survey and the ADHD Self Report Scale. After that, participants took the final rank order retention test, followed by the last ventilator concept test. At the end of the study, participants were thanked and Qualtrics automatically assigned class credit as compensation.

Results

Description of the Sample

Table 2 contains a description of the participant sample used in this study.

Table 2

Demographics

General	*n*=43 Control		*n*=33 smart DR		*n*=41 Faded DR		*n*=117 All Participants	
	M	SD	M	SD	M	SD	M	SD
ASRS[1]	51.91	8.39	49.97	10.60	57.54	10.31	53.33	10.17
ACS[2]	60.95	7.03	61.73	5.84	56.63	7.33	59.66	7.13
Retention 1 Completion Time[3]	172.9	83.37	220.6	116.9	195.3	86.4	194.2	96.1
Transfer Completion Time[3]	136.5	79.7	175.5	109.3	163.6	85.9	157	91.7
Frequency								
Age	18 to 24=42 25 to 34=1		18 to 24=31 25 to 34=2		18 to 24=40 25 to 34=1		18 to 24=113 25 to 34=4	
Gender[4]	17M 26F 0NB		17M 16F 0NB		13M 26F 2NB		47M 68F 2NB	
English as a First Language	Yes=43 No=0		Yes=31 No=2		Yes=41 No=1		Yes=115 No=3	
VR Experience	None=32 Some=11 A lot=0		None=18 Some=15 A lot=0		None=27 Some=14 A lot=0		None=77 Some=40 A lot=0	
Knowledge of Medical Ventilation Procedures	Excellent=0 Good=0 Average=0 Poor=14 None=29		Excellent=0 Good=0 Average=0 Poor=8 None=25		Excellent=0 Good=1 Average=1 Poor=7 None=32		Excellent=0 Good=1 Average=1 Poor=29 None=86	

**Note.* [1]Adult ADHD Self Report Scale (Kessler et al., 2005) and the [2]Attentional Control Scale (Derryberry & Reed, 2002). [3]Measured in seconds. [4]Male, Female, Non-Binary.

Performance on Retention and Transfer

Retention Test Results

Pure Ranking.

A two sample t-test was conducted to compare pure retention scores between the faded and control conditions, as well as the faded and smart DR conditions. For faded and control, there was no significant difference between the groups $t(82) = -.017$, $p = 1$. Using Cohen's *d*, a negative effect size of -.003 was found. To assess differences between the faded and smart DR group, we ran a two sample t-test and found no significant difference, $t(72) = 1.8$, $p = .07$. There was a small effect size as calculated by Cohen's *d* (*d*=.4).

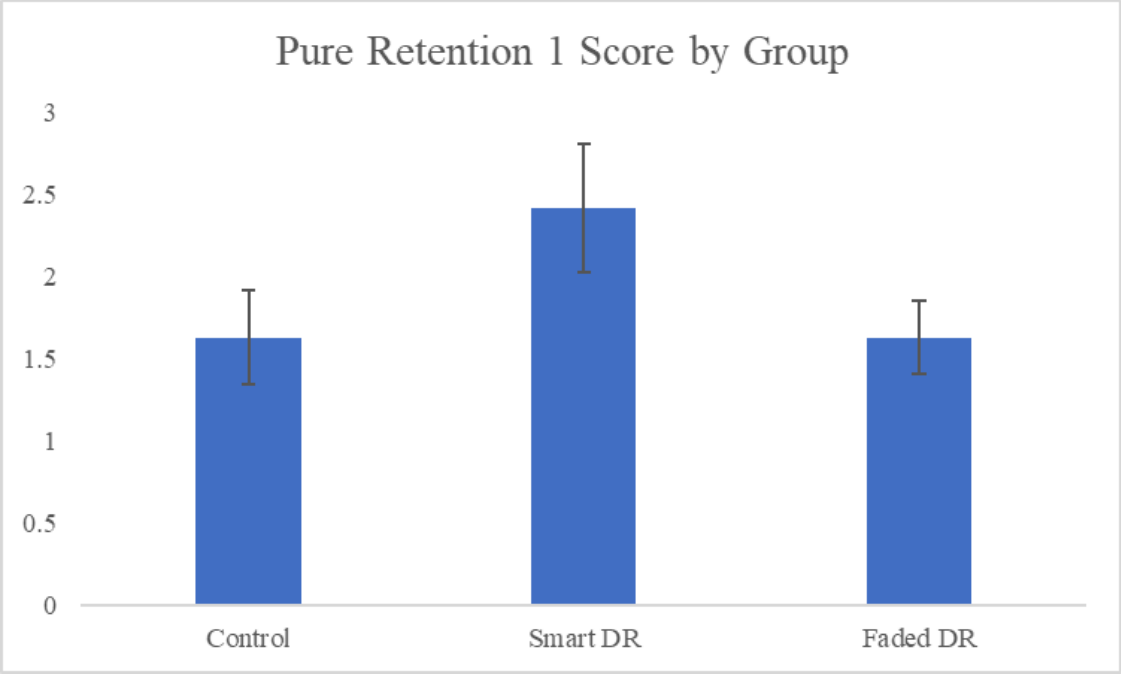


Figure 8. Mean scores for the first retention test separated by condition. Error bars represent standard errors of the mean (SEM).

Distance Score.

A two sample t-test was conducted to compare retention distance scores between the faded and control conditions, as well as the faded and smart DR conditions. For faded and control, there was no significant difference, $t(82) = 1.5$, $p = .1$, with a small effect size of .333 as calculated by Cohen's *d*. For differences between smart-DR and faded, we also ran a two sample t-test and found no significant difference, $t(72) = -.52$, $p = .6$, with a small effect size (d=-.12).

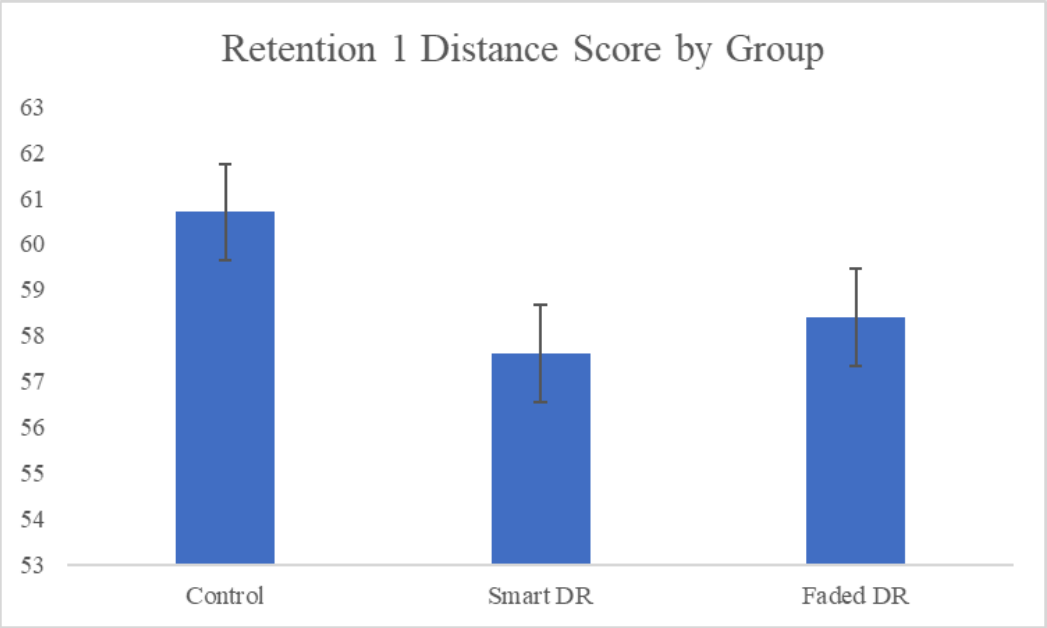


Figure 9. Mean scores for distance score on the first retention test separated by condition. Error bars represent standard errors of the mean (SEM).

Longest Sequence.

A two sample t-test was conducted to compare retention longest sequence scores between the faded and control conditions, as well as the faded and smart DR conditions. For faded and control, there was no significant difference between the groups $t(82) = -1.0$, $p = .3$. There was a small effect size of -.228 as calculated by Cohen's *d*. For differences between faded and smart DR, there was no significant difference, $t(72) = 1.9$, $p = .06$. There was a small effect size of .44.

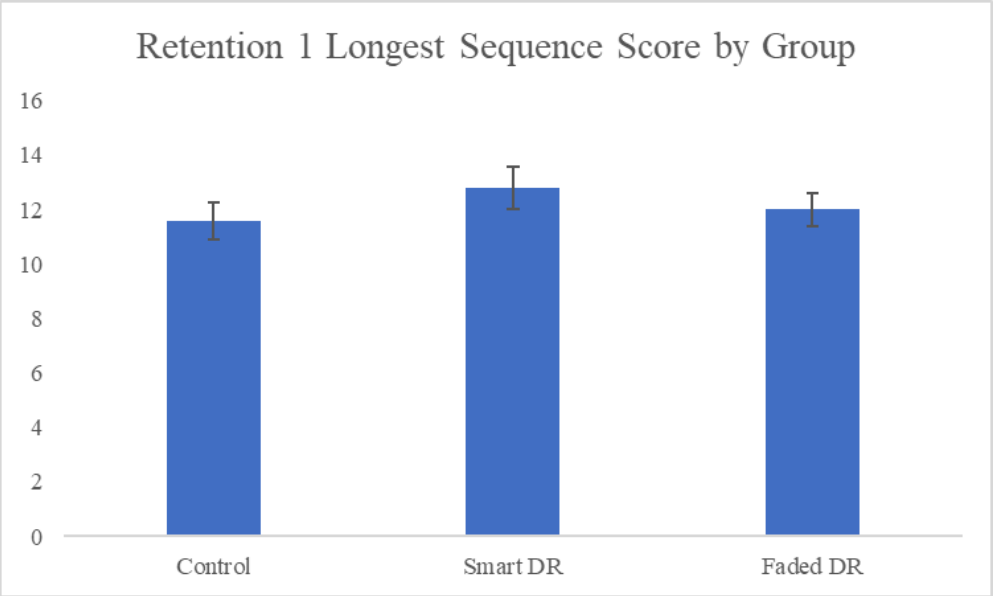


Figure 10. Mean scores for longest sequence on the first retention test separated by condition. Error bars represent standard errors of the mean (SEM).

Completion Time.

To assess group differences in completion time on the first retention test, a two sample t-test was conducted. There was no significant difference between the control and faded group, $t(82) = -1.2$, $p = .2$, with a negative effect size of -.26 as calculated by Cohen's D. For differences between smart DR and faded, there was no significant difference, $t(72) = 1.1$, $p = .3$, with a small effect size of .25.

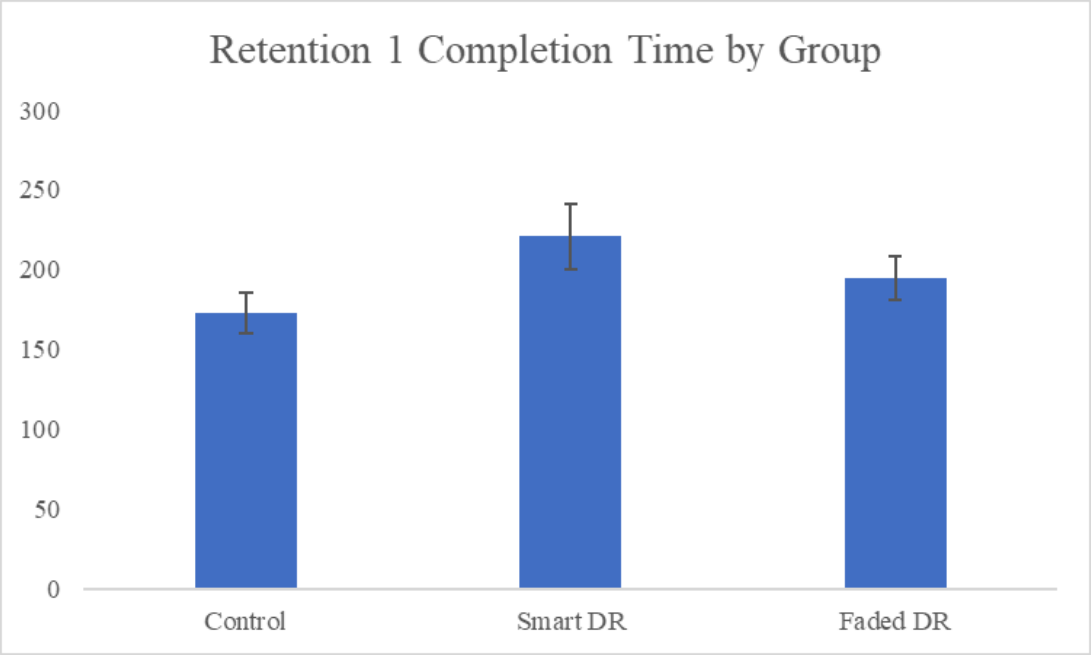


Figure 11. Mean scores for completion time on the first retention test separated by condition. Error bars represent standard errors of the mean (SEM).

Transfer Test Results

Pure Transfer Ranking.

To determine group differences on pure transfer score, two sample t-tests were conducted. There was no difference between the faded and control group, $t(82) = -.69$, $p = .5$, with a small effect size of -.15 as calculated by Cohen's *d.* There was also no significant difference between faded and smart DR, $t(72) = .07$, $p = .9$, with a small effect size of .02,

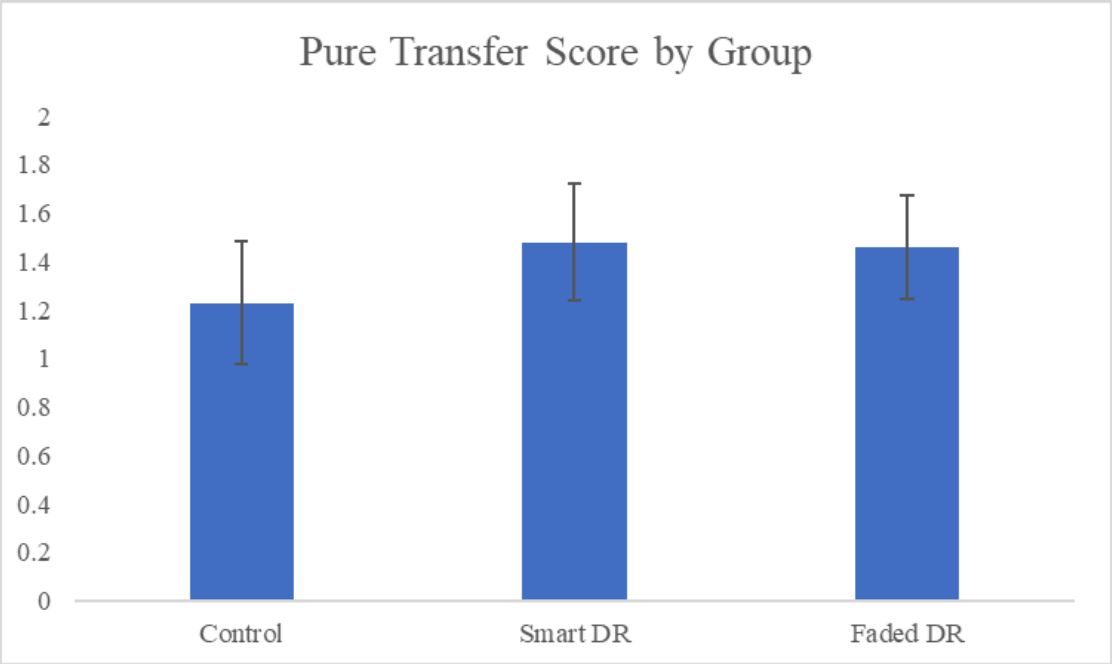


Figure 12. Mean scores for the transfer ranking test separated by condition. Error bars represent standard errors of the mean (SEM).

Distance Score.

For distance score, two sample t-tests were conducted to determine if type of DR had an effect on performance. The results showed that there was no significant difference between control and faded $t(82) = .35$, $p = .7$, with Cohen's d calculated as a small effect size (d=.07). For faded compared to smart DR, there was no significant difference, $t(72) = -.09$, $p = .3$, with a small effect size of -.22.

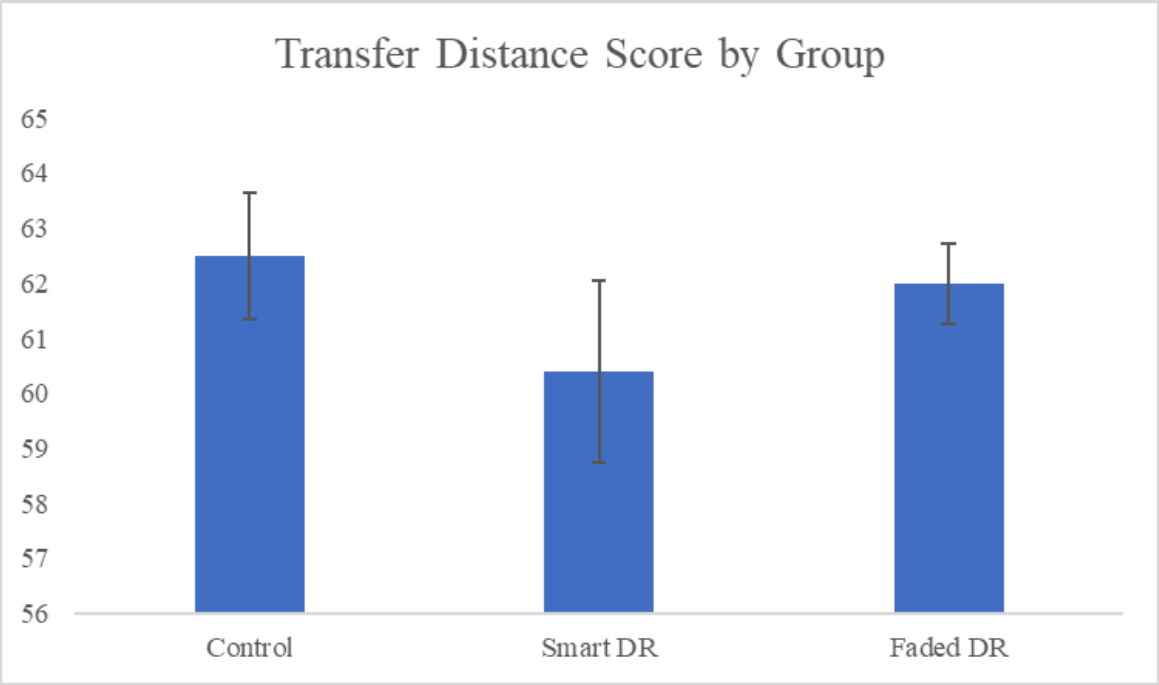


Figure 13. Mean scores for distance for the transfer test, by condition. Error bars represent standard errors of the mean (SEM).

Longest Sequence.

To determine if type of DR affected performance, two sample t-tests were conducted. There was no significant difference between control and faded $t(82) = -.2$, $p = .8$. There was a small effect size (d=-.04). For faded and smart DR, there was no significant difference between groups, $t(72) = 1.3$, $p = .2$, with a small effect size of .29.

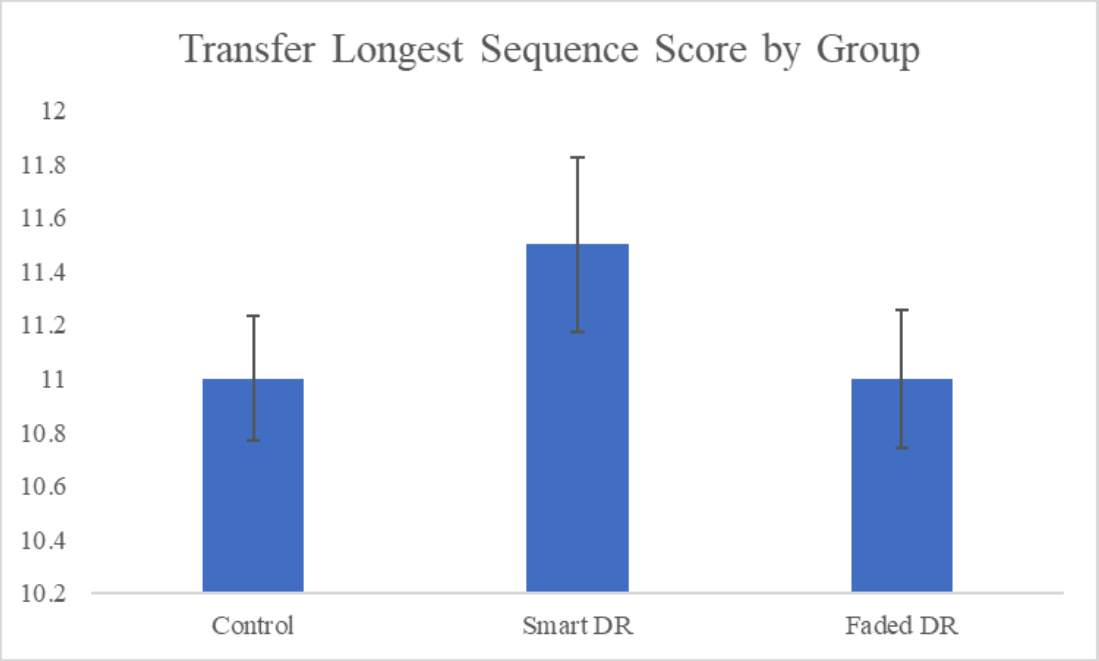


Figure 14. Mean scores for longest sequence on the transfer test, by condition. Error bars represent standard errors of the mean (SEM).

Completion Time.

To determine differences by condition on completion time, two sample t-tests were conducted. There was no significant difference between faded and control $t(82) = -1.5$, $p = .1$, with a small effect size (d=-.3) For completion time between faded and smart DR, there was no significant difference between groups, $t(72) = .51$, $p = .6$. There was a small effect size of .12.

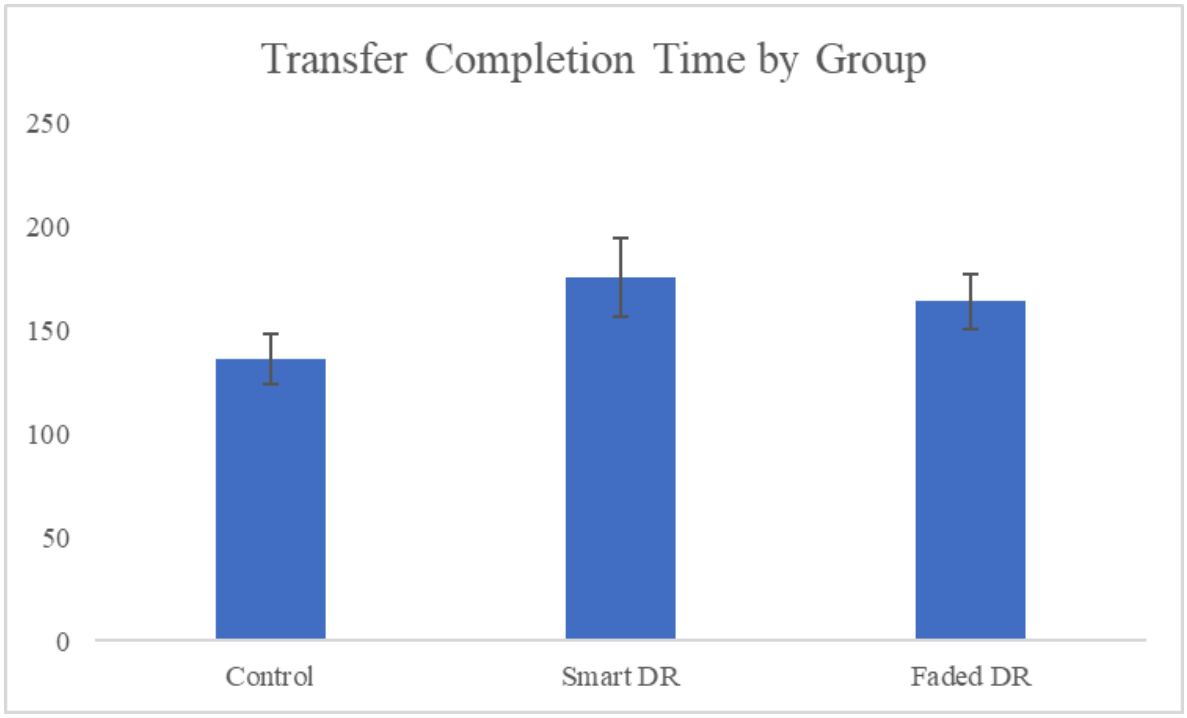


Figure 15. Mean completion time on transfer test separated by condition. Error bars represent standard errors of the mean (SEM).

Performance on Concept Tests

Change in accuracy was assessed between retention 1 and retention 2 concept tests. Questions where the mean change was higher than 6% were included in the retention concept test composite score, as they indicated learning from pre- to post-test. Questions 3, 4, 5, 6, and 7 were included in the composite score, and all questions included in the test can be seen in Appendix D.

For the transfer concept test composite score, questions where the mean accuracy was higher than 25% (chance of getting the correct answer) were included into the composite score. Questions 2, 3, 5, 8, and 9 were included in the composite score, and the full test can be seen in Appendix E.

Retention Concept Test Results.

For the composite concept test score, two sample t-tests were used. There was no significant difference between control and faded, $t(80) = -.2$, $p = .8$. Cohen's d was calculated as -.04. There was no significant difference between faded and smart DR, $t(70) = -.44$, $p = .7$, with a small effect size of .10.

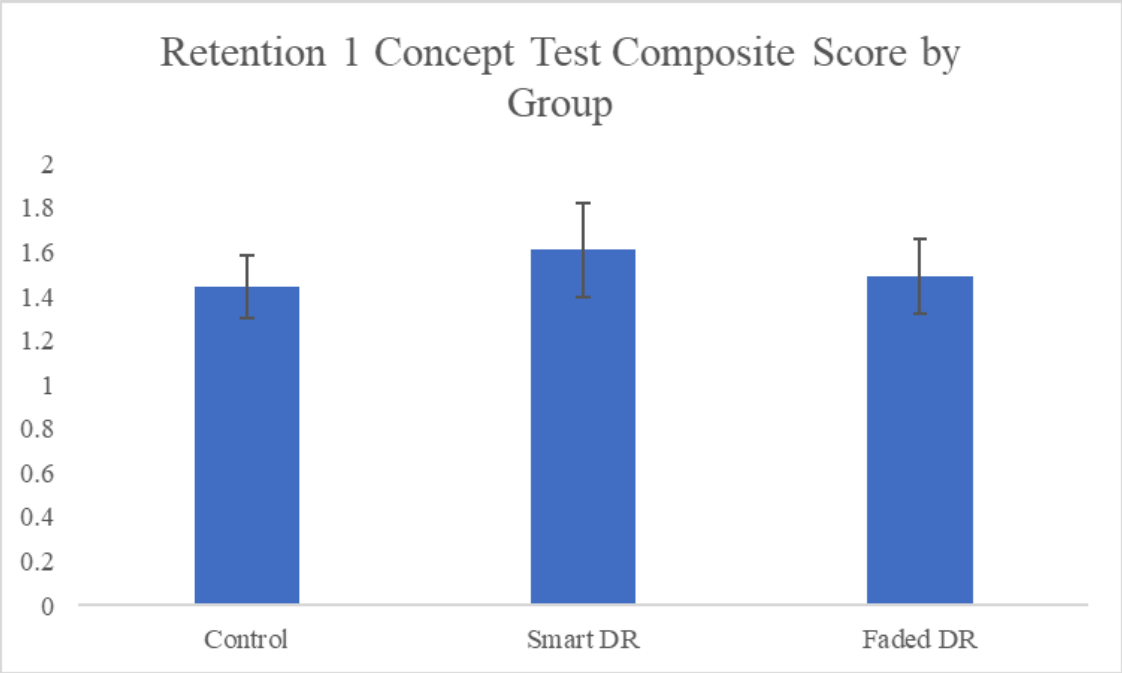


Figure 16. Mean scores for the retention concept test separated by condition. Error bars represent standard errors of the mean (SEM).

Transfer Concept Test Results.

We conducted two sample t-tests to determine group differences in concept test performance. For control and faded, there was no significant difference, $t(80) = .43$, $p = .7$, with a small effect size ($d = .09$). For smart DR and faded, there was no significant difference between groups, $t(69) = -1.6$, $p = .1$. Cohen's d was calculated as -.4.

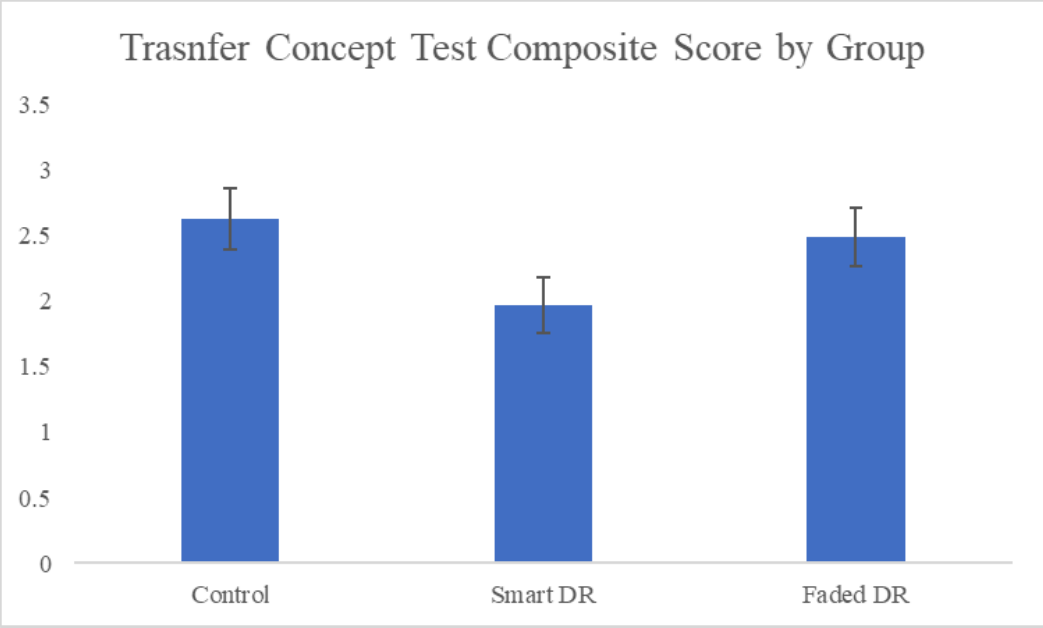


Figure 17. Mean scores the transfer concept test separated by condition. Error bars represent standard errors of the mean (SEM).

Effects of Individual Differences

Correlations were run between Adult ADHD scores, ACS scores, and the outcome measures. There were no significant correlations between the attentional measures and any of the outcome variables.

Table 3
Correlations Between Attentional Measures and Outcome Measures

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11
1. ACS	0.07	0.34											
2. ASRS	0.02	0.36	-.72** [-.91, -.24]										
3. Pure Retention	0.18	0.35	.33 [-.30, .76]	-.54 [-.85, .05]									
4. Distance, Retention	-0.04	0.39	-.22 [-.71, .40]	.45 [-.17, .81]	-.79** [-.94, -.40]								
5. Longest Sequence, Retention	0.19	0.39	.35 [-.28, .77]	-.55 [-.85, .04]	.78** [.36, .93]	-.84** [-.95, -.51]							
6. Retention Completion Time	0.18	0.38	.06 [-.53, .62]	-.44 [-.81, .18]	.51 [-.10, .84]	-.70* [-.91, -.21]	.62* [.08, .88]						
7. Pure Transfer	0.18	0.32	.34 [-.29, .77]	-.32 [-.75, .32]	.42 [-.20, .80]	-.38 [-.78, .25]	.58* [.00, .86]	.36 [-.27, .78]					
8. Distance Score, Transfer	-0.08	0.39	-.38 [-.78, .25]	.33 [-.30, .76]	-.46 [-.82, .15]	.43 [-.19, .81]	-.65* [-.89, -.13]	-.62* [-.88, -.07]	-.70* [-.91, -.22]				
9. Longest Sequence, Transfer	0.18	0.39	.38 [-.25, .78]	-.51 [-.84, .09]	.62* [.08, .88]	-.56 [-.86, .02]	.79** [.40, .94]	.53 [-.06, .85]	.69* [.18, .90]	-.86** [-.96, -.56]			
10. Transfer Completion Time	0.16	0.40	.11 [-.50, .64]	-.50 [-.83, .11]	.52 [-.08, .84]	-.74** [-.92, -.28]	.65* [.12, .89]	.94** [.80, .98]	.32 [-.31, .75]	-.61* [-.88, -.06]	.57 [-.01, .86]		
11. Retention Concept Test Composite Score	0.16	0.28	-.26 [-.72, .37]	.18 [-.44, .68]	.18 [-.44, .68]	-.20 [-.70, .42]	.09 [-.51, .63]	.13 [-.48, .65]	.14 [-.48, .66]	-.16 [-.67, .45]	.07 [-.53, .62]	.12 [-.49, .65]	
12. Transfer Concept Test Composite Score	0.14	0.28	-.30 [-.75, .33]	.31 [-.32, .75]	-.22 [-.70, .41]	.12 [-.48, .65]	-.16 [-.67, .46]	.02 [-.56, .59]	-.11 [-.64, .50]	-.03 [-.59, .55]	-.14 [-.66, .47]	-.07 [-.62, .52]	.29 [-.34, .74]

Means, standard deviations, and correlations with confidence intervals

Note. M and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates $p < .05$. ** indicates $p < .01$.

Additionally, to test whether those with lower attentional abilities received more of a benefit by being in the faded DR group, we looked at main effects for Adult ADHD scores and ACS scores on the retention and transfer tests using multivariate linear regression analyses. Only distance and longest sequence scores were used, as they were more normally distributed than other outcome variables. Out of the 4 models, only the model looking at the interaction between attentional measures and transfer longest distance scores were significant.

Model 1, which tested whether there was a relationship between ACS or ADHD scores and performance as measured by retention distance scores, was not significant $F(11, 105) = .991$, $p = .46$, R2 = .09). Additionally, we tested if those with higher ADHD scores received more of a performance benefit in the faded group, but found no significant relationship ($\beta = .39$, $p = .74$). There was also no significant relationship between being in the faded group and low ACS scores ($\beta = .29$, $p = .79$)

Model 2, which tested whether there was a relationship between ACS or ADHD scores and performance as measured by retention longest sequence scores, was significant, $F(11, 105) = 2.1$, $p = .02$, R^2=.18), however, the model only explained 18% of the variance in the sample. When looking at the interaction between high ADHD scores and being in the faded group, there was no significant relationship ($\beta = .55$, $p = .07$). There was also no significant relationship between being in the faded group and low ACS scores ($\beta = .55$, $p = .06$)

Model 3 which tested whether there was a relationship between ACS or ADHD scores and performance as measured by transfer distance scores was not significant $F(11, 105) = .89$, $p = .55$, $R^2 = .08$). Furthermore, when looking at the interaction between high ADHD scores and being in the faded group, there was no significant relationship found ($\beta = -1.5$, $p = .23$). There

was also no significant relationship between being in the faded group and low ACS scores (β = -1.2, p = .31).

Model 4 which tested whether there was a relationship between ACS or ADHD scores and performance as measured by transfer longest distance was not significant $F(11, 105) = .11$, $p = .4$, $R^2 = .10$). We also tested if those with higher ADHD scores received more of a performance benefit in the faded group, but found no significant relationship (β = .37, p = .2). There was also no significant relationship between being in the faded group and low ACS scores (β = .35, p = .2)

Discussion

Overview

The central aim of this study was to understand how controlling distraction via DR videos can affect training, specifically retention and near transfer. Additionally, the study was designed to investigate how individual differences play a role in one's learning of a complex task, and how the type of DR used may help overcome these differences. However, in all conditions, few learned the concepts needed for the processes of assembling the ventilator, nor did any condition transfer their knowledge to a similar system in a new domain.

Difficulties, in the form of distractions, may allow for better retention and transfer, when the main task is likely to contain distractions. However, experiencing a highly distracting environment during initial learning likely impedes learning (Bjork & Bjork, 2020) by overwhelming the learner (Zepeda et al., 2020). In the current experiments, the task learners were trained on was a hard task to complete in general, even in conditions where distractions were removed. The theory of transfer appropriate processing would predict that for a task such as learning the steps for ventilator assembly, learners should practice retrieving information during

learning (Dobson & Linderholm, 2015). Examination of the data from the main study suggests they were unable to do so, instead passively reviewing the videos.

Transfer performance was also low across all conditions. Transfer appropriate processing, when successful, occurs when the processes engaged during training and those engaged during tests of retention overlap (Franks et al., 2000). However, an understanding of the target retention skill is likely necessary for transfer to occur, as transfer requires the learner to understand the underlying concepts of the target task (Kim & Lee, 2001). Unfortunately, in the current study, participants were unable to learn the target task, and which likely impacted their performance on the transfer task. Additionally, there may be use in training to automaticity, rather than to proficiency to achieve transfer. In our study, we attempted to train users to proficiency, as measured by tests of retention. A recent study used automaticity (with the addition of a second task) as a measure of learning for skill transfer between simulator laparoscopic surgery and the clinical environment (Stefanidis et al., 2012). While there was a second task in our study (ignoring distractions), the repetition in Stefanidis et al. 2012 was based on achieving expert level performance; participants trained until they achieved expert level performance on the primary and the second task. Participants in the current study did not have a chance to practice more than once, or become experts in the ventilator assembly, before being presented with the second task of ignoring distractions, which may explain why their performance on a transfer task, as well as retention was low.

When considering the study in the context of Cognitive Load Theory (CLT), we proposed that distractions may be intrinsic load, especially if those distractions occur during real world tests of retention. In our study, distractions were intrinsic load, even though they were irrelevant to the task. Generally, task irrelevant distractors may not enter working memory, but someone

new to a task may not immediately know when something is task relevant or task irrelevant. (Hakim et al., 2021). Experts are generally better at this skill, which has also been called "information reduction" or "selective attention" (Lavie et al., 2004; Szulewski et al., 2021). In our study, participants did not have the chance to become experts, so they may not have been able to discriminate between task relevant and irrelevant distractions. While we did tell participants ahead of time that they would be asked about the task steps, that prior knowledge may not have been enough to determine which auditory or visual elements were irrelevant. Furthermore, even in the smart DR condition, with less distractions to classify as relevant or irrelevant, participants performed poorly. The smart DR condition may have been distracting to participants, as there were still auditory and visual distractions occurring in the environment that were unrelated to the ventilator assembly.

Additionally, there are some learning task characteristics based on CLT that may help reduce cognitive load when learning: the split attention effect and the worked example effect (Paas & van Merriënboer, 2020).

The split attention effect states that information from one source, rather than across multiple sources (spatial or temporal) imposes a lower extraneous load. In the current study, there were two sources of information: the steps being said out loud in the video, and the screen with instructions. However, the screen was not in the video the entire time; the video panned around the room as different pieces were located. Even so, with just the steps being said aloud, new learners may not have known that the voice was the sole source of information, and may have attended to the irrelevant auditory information (Hakim et al., 2021). This may have increased their cognitive load, and may explain the poor performance.

The other learning task characteristic, the worked example effect, states that a novice learns more from examples that provide them with the solution versus solving the problems themselves, as problem solving introduces a high cognitive load and leaves no resources for learning (Paas & van Merriënboer, 2020). In this study, participants did receive a worked example: the solution to the ventilator assembly task, presented via auditory steps and visual steps. Despite this, performance was low. This may have been due to the high extraneous load of distractions, which would leave no resources in working memory for learning the actual task. Additionally, the length of time spent with the ventilator solution was limited to 3-4 minutes.

Emphasis change and scaffolding may also be an effective method of reducing cognitive load, initially (Gopher et al., 2007). However, the task switching required in emphasis change may actually increase cognitive load. And while scaffolding gradually increases the amount of load, if learners never adapted to the initial load, they may still perform poorly. Ideally in scaffolding, instructional aids become redundant as knowledge increases (Herbert et al., 2021). However in our study, learners did not gain more knowledge, so any instructional aid would have been necessary, not redundant. This may explain the poor performance in the current study.

Learning via video is just one of many methods that medical practitioners use. Students typically use multiple methods of learning to achieve learning goals, such as textbooks, hands on learning, student-led lectures, traditional lectures, all of which can take place in person or remotely. In comparison to in person learning, cognitive engagement may be higher during asynchronous learning. (Thacker et al., 2022). The current study utilized video training methods that can be delivered asynchronously, and while they did not lead to learning, they may be used in conjunction with other learning methods. This may be especially useful in cases where AR/VR or DR technologies are not feasible to use as the sole training method.

Limitations

Some of the limitations in this study were due to remote data collection. To reduce the chance of participants skipping or rewatching the videos, we utilized Qualtrics' timing to auto advance only once the video was over. However, controlling external distractions unrelated to the study was not possible. Additionally, the interval where participants completed the transfer test, Adult ADHD measure and the medical knowledge was based on participants' speed in filling them out and may have varied between participants. Furthermore, whether participants made serious efforts to remember the information from the videos was unable to be controlled. There is also a possibility that fading distractions has no effect on learners, even with fully engaged participants and no external distractions.

Additionally, the amount of analyses that were run may have increased the possibility of a Type I Error. In future studies, this may be mitigated by increasing the sample size. Finally, the pure retention and transfer scores were both highly skewed, even when transforming the data. The low performance on the pure scores impacted the distance and longest sequence scores, despite those measures being more forgiving.

Future Research

Future training initiatives could involve automated DR, where learners can simply choose their level of difficulty, and have the DR system provide a level of faded-DR that matches their skill level. Being able to choose an appropriate difficulty level can motivate learners to cope with and embrace difficulties during training (Zepeda et al., 2020). An example would be allowing learners to determine how many distractions they experience during the training session.

However, this may result in learners over-relying on DR, which may lead to detriments in performance when not able to limit distractions.

Additionally, in this study, participants had no reason to believe that the DR system would fail to filter out needed information or diminish the environment incorrectly. However, in future applications where DR is automated, the effect of one's trust in automation becomes apparent. Users with lower levels of trust in automation may pause to verify information provided by an automated system, which can increase workload and undermine the usefulness of the system (St. John et al., 2005). These factors can impact how much a user will rely on DR during complex tasks. Furthermore, levels of trust are dependent on the reliability of the system, how the automation is displayed, difficulty of the task, and the consequences of misses and false alarms; all of which are moderated by individual differences (Parasuraman & Riley, 1997).

Another avenue of research is to revisit the Desktop VR system used in the pilot study. While learning via video is just one of many methods that medical practitioners use, students typically use multiple methods of learning to achieve learning goals, such as textbooks, traditional lectures, student-led lectures, and hands-on learning, all of which can take place in person or remotely. With the Desktop VR system, learners may benefit from hands-on learning remotely. Specifically, the interactivity of Desktop VR makes learning a more active process, which can lead to higher performance on knowledge tests (El Sadik & Al Abdulmonem, 2021).

Conclusion

AR and other mixed reality technologies are already being used for learning and training purposes in the education, defense, and manufacturing industries, especially as the cost of accessing this technology lowers. However, the concept of specifically using DR for learning is

newer, as DR has primarily been used in computer science fields to study real-time object removal in fields such as interior design.

Ultimately, there are still unanswered questions about the use of DR for learning step-based tasks under distraction. More research is needed to integrate theories of learning with DR successfully. While this study did not show that distractions and emphasis change delivered via DR can aid retention and transfer, there still may be a use in diminishing distractions for learners. Even successful learners may still need to develop the skills needed to ignore distractions and DR may be able to help. Combining novel DR applications and effective training methods that utilize it will allow for new approaches to learning, and as mixed reality technology becomes more accessible, others can benefit from the use of this type of research.

References

Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. *Psychological Bulletin*, *101*(1), 41–74. https://doi.org/10.1037/0033-2909.101.1.41

Aoyagi, Y., Ohnishi, E., Yamamoto, Y., Kado, N., Suzuki, T., Ohnishi, H., Hokimoto, N., & Fukaya, N. (2019). Feedback protocol of 'fading knowledge of results' is effective for prolonging motor learning retention. *Journal of Physical Therapy Science*, *31*(8), 687–691. https://doi.org/10.1589/jpts.31.687

Azuma, R. T. (1997). A survey of augmented reality. *PRESENCE: Teleoperators & Virtual Environments*, *6*(4), 355. https://doi.org/10.1162/pres.1997.6.4.355

Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition*, *9*(4), 475–479. https://doi.org/10.1016/j.jarmac.2020.09.003

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 657-668.

Bonwell, C. C., & Eison, J. A. (1991). *Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports*. ERIC Clearinghouse on Higher Education, The George Washington University, One Dupont Circle, Suite 630, Washington, DC 20036-1183 ($17. https://eric.ed.gov/?id=ED336049

Boot, W. R., Basak, C., Erickson, K. I., Neider, M., Simons, D. J., Fabiani, M., Gratton, G., Voss, M. W., Prakash, R., Lee, H., Low, K. A., & Kramer, A. F. (2010). Transfer of skill engendered by complex task training under conditions of variable priority. *Acta Psychologica*, *135*(3), 349–357. https://doi.org/10.1016/j.actpsy.2010.09.005

Brevik, E. J., Lundervold, A. J., Haavik, J., & Posserud, M.-B. (2020). Validity and accuracy of the Adult Attention-Deficit/Hyperactivity Disorder (ADHD) Self-Report Scale (ASRS) and the Wender Utah Rating Scale (WURS) symptom checklists in discriminating between adults with and without ADHD. *Brain and Behavior*, *10*(6), e01605. https://doi.org/10.1002/brb3.1605

Chucherd, O., Vallibhakara, S. A.-O., Paiwattananupant, K., Puranitee, P., Wattanayingcharoenchai, R., & Vallibhakara, O. (2023). The effect of online video-assisted teaching program on medical students learning procedure of fractional curettage. *BMC Medical Education*, *23*, 82. https://doi.org/10.1186/s12909-023-04052-3

Clapp, W. C., Rubens, M. T., & Gazzaley, A. (2010). Mechanisms of working memory disruption by external interference. *Cerebral Cortex*, *20*(4), 859–872. https://doi.org/10.1093/cercor/bhp150

Derryberry, D., & Reed, M. A. (2002). Anxiety-related attentional biases and their regulation by attentional control. *Journal of Abnormal Psychology*, *111*(2), Article 2. https://doi.org/10.1037/0021-843X.111.2.225

Detterman, D. K., & Sternberg, R. J. (1993). *Transfer on trial: Intelligence, cognition, and instruction.* Ablex Publishing.

Dobson, J. L., & Linderholm, T. (2015). The effect of selected "desirable difficulties" on the ability to recall anatomy information. *Anatomical Sciences Education*, *8*(5), 395–403. https://doi.org/10.1002/ase.1489

Eckert, M., Volmerg, J. S., & Friedrich, C. M. (2019). Augmented Reality in Medicine: Systematic and Bibliographic Review. *JMIR mHealth and uHealth*, *7*(4), e10967. https://doi.org/10.2196/10967

El Sadik, A., & Al Abdulmonem, W. (2021). Improvement in Student Performance and Perceptions through a Flipped Anatomy Classroom: Shifting from Passive Traditional to Active Blended Learning. *Anatomical Sciences Education*, *14*(4), 482–490. https://doi.org/10.1002/ase.2015

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology*, *18*(3), Article 3. https://doi.org/10.1016/0010-0285(86)90001-0

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Forster, S., & Lavie, N. (2008). Failures to ignore entirely irrelevant distractors: The role of load. *Journal of Experimental Psychology: Applied*, *14*(1), 73–83. https://doi.org/10.1037/1076-898X.14.1.73

Franks, J. J., Bilbrey, C. W., Lien, K. G., & McNamara, T. P. (2000). Transfer-appropriate processing (TAP) and repetition priming. *Memory & Cognition*, *28*(7), 1140–1151. https://doi.org/10.3758/BF03211815

Gopher, D., Kramer, A., Wiegmann, D., & Kirlik, A. (2007). Emphasis change as a training protocol for high-demand tasks. *Attention: From Theory to Practice*, *101*, 209–224.

Hambrick, D. Z., Oswald, F. L., Darowski, E. S., Rench, T. A., & Brou, R. (2010). Predictors of multitasking performance in a synthetic work paradigm. *Applied Cognitive Psychology*, *24*(8), 1149–1167. https://doi.org/10.1002/acp.1624

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology* (Vol. 52, pp. 139–183). North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9

Herbert, B., Wigley, G., Ens, B., & Billinghurst, M. (2022). Cognitive load considerations for augmented reality in network security training. *Computers & Graphics*, *102*, 566–591. https://doi.org/10.1016/j.cag.2021.09.001

Hicks, W. B., Pryor, M., Sprufera, J., Greene, A., Byrne, V., & McLaughlin, A. C. (2018). Time pressure and teams: contextual factors influencing task experience when using a cognitive aid. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 125–129. https://doi.org/10.1177/1541931218621029

Hoff, K. A., & Bashir, M. (2015). Trust in automation: integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Hoonakker, P., Carayon, P., Gurses, A., Brown, R., McGuire, K., Khunlertkit, A., & Walker, J. M. (2011). Measuring workload of ICU nurses with a questionnaire survey: The nasa task load index (TLX). *IIE Transactions on Healthcare Systems Engineering*, *1*(2), 131–143. https://doi.org/10.1080/19488300.2011.609524

Hughes, M. A., Swan, L., Taylor, C. L., Ilin, R., Partridge, R., & Brennan, P. M. (2022). The Impact of Novel Nontechnical Stressors (Visual and Auditory) on Simulated Laparoscopic Task Performance Among Surgeons and Students. *Journal of Laparoendoscopic & Advanced Surgical Techniques*, *32*(2), Article 2. https://doi.org/10.1089/lap.2021.0695

Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews. Neuroscience*, *2*(3), 194–203. http://dx.doi.org.prox.lib.ncsu.edu/10.1038/35058500

Judah, M. R., Grant, D. M., Mills, A. C., & Lechner, W. V. (2014). Factor structure and validation of the attentional control scale. *Cognition & Emotion*, *28*(3), 433–451.

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., Howes, M. J., Jin, R., Secnik, K., Spencer, T., Ustun, T. B., & Walters, E. E. (2005). The World Health Organization adult ADHD self-report scale (ASRS): A short screening scale for use in the general population. *Psychological Medicine*, *35*(2), 245–256. http://dx.doi.org.prox.lib.ncsu.edu/10.1017/S0033291704002892

Kim, J. H., & Lee, C. (2001). Implications of near and far transfer of training on structured on-the-Job training. *Advances in Developing Human Resources*, *3*(4), 442–451. https://doi.org/10.1177/15234220122238508

Krakauer, J. W., Ghez, C., & Ghilardi, M. F. (2005). Adaptation to visuomotor transformations: consolidation, interference, and forgetting. *Journal of Neuroscience*, *25*(2), 473–478. https://doi.org/10.1523/JNEUROSCI.4218-04.2005

Kramer, A. F., Larish, J. F., & Strayer, D. L. (1995). Training for attentional control in dual task settings: A comparison of young and old adults. *Journal of Experimental Psychology: Applied*, *1*(1), 50–76. https://doi.org/10.1037/1076-898X.1.1.50

Lai, A.-F., Chen, C.-H., & Lee, G.-Y. (2019). An augmented reality-based learning approach to enhancing students' science reading performances from the perspective of the cognitive load theory. *British Journal of Educational Technology*, *50*(1), 232–247. https://doi.org/10.1111/bjet.12716

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load Theory of Selective Attention and Cognitive Control. *Journal of Experimental Psychology: General*, *133*(3), 339–354. https://doi.org/10.1037/0096-3445.133.3.339

Lazaro, M. J., Kang, Y., Yun, M. H., & Kim, S. (2021). The effects of visual complexity and decluttering methods on visual search and target detection in cockpit displays.

International Journal of Human–Computer Interaction, *37*(7), 588–600. https://doi.org/10.1080/10447318.2021.1890491

Le Chénéchal, M., & Chatel-Goldman, J. (2018). HTC Vive Pro time performance benchmark for scientific research. *ICAT-EGVE 2018*.

Lee, J., Kim, H., Kim, K. H., Jung, D., Jowsey, T., & Webster, C. S. (2020). Effective virtual patient simulators for medical communication training: A systematic review. *Medical Education*, *54*(9), Article 9. https://doi.org/10.1111/medu.14152

Lefor, A. K., Harada, K., Kawahira, H., & Mitsuishi, M. (2020). The effect of simulator fidelity on procedure skill training: A literature review. *International Journal of Medical Education*, *11*, 97–106. https://doi.org/10.5116/ijme.5ea6.ae73

Morgan, P. J., Cleave-Hogg, D., McIlroy, J., & Devitt, J. H. (2002). Simulation technology: A comparison of experiential and visual learning for undergraduate medical students. *Anesthesiology*, *96*(1), Article 1. https://doi.org/10.1097/00000542-200201000-00008

Murali, S., Paul, K. D., McGwin, G., & Ponce, B. A. (2021). Updates to the Current Landscape of Augmented Reality in Medicine. Cureus, 13(5), e15054.

Murph, I., McDonald, M., Richardson, K., Wilkinson, M., Robertson, S., Karunakaran, A., Gandy Coleman, M., Byrne, V., & McLaughlin, A. C. (2021). Diminishing reality: potential benefits and risks. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *65*(1), 164–168. https://doi.org/10.1177/1071181321651103

O'Hara, S., Dwyer, N., & Simon, M. (2007). Intelligent agents for decluttering command-and-control displays. *2007 International Conference on Integration of Knowledge Intensive Multi-Agent Systems*, 230–235. https://doi.org/10.1109/KIMAS.2007.369814

Paas, F., & van Merriënboer, J. J. G. (2020). Cognitive-Load Theory: Methods to Manage Working Memory Load in the Learning of Complex Tasks. *Current Directions in Psychological Science*, *29*(4), 394–398. https://doi.org/10.1177/0963721420922183

Papanikolaou, I. G., Haidopoulos, D., Paschopoulos, M., Chatzipapas, I., Loutradis, D., & Vlahos, N. F. (2019). Changing the way we train surgeons in the 21th century: A narrative comparative review focused on box trainers and virtual reality simulators. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, *235*, 13–18. https://doi.org/10.1016/j.ejogrb.2019.01.016

Parasuraman, R., & Riley, V. (1997). Humans and automation: use, misuse, disuse, abuse. *Human Factors*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Perkins, D. N., & Salomon, G. (1992). Transfer of learning. *International Encyclopedia of Education*, *2*, 6452–6457.

Ragan, E., Wilkes, C., Bowman, D. A., & Hollerer, T. (2009). Simulation of augmented reality systems in purely virtual environments. *2009 IEEE Virtual Reality Conference*, 287–288. https://doi.org/10.1109/VR.2009.4811058

Redick, T. S. (2016). On the relation of working memory and multitasking: memory span and synthetic work performance. *Journal of Applied Research in Memory and Cognition*, *5*(4), 401–409. https://doi.org/10.1016/j.jarmac.2016.05.003

Sala, G., Aksayli, N. D., Tatlidil, K. S., Tatsumi, T., Gondo, Y., & Gobet, F. (2019). Near and far transfer in cognitive training: a second-order meta-analysis. *Collabra: Psychology*, *5*(1), 1-22. https://doi.org/10.1525/collabra.203

Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin*, *95*(3), 355–386. https://doi.org/10.1037/0033-2909.95.3.355

Sandrone, S., Scott, G., Anderson, W. J., & Musunuru, K. (2021). Active learning-based STEM education for in-person and online learning. *Cell*, *184*(6), Article 6. https://doi.org/10.1016/j.cell.2021.01.045

Schmidt, R. A., & Bjork, R. A. (1992). New Conceptualizations of practice: common principles in three paradigms suggest new concepts for training. *Psychological Science*, *3*(4), 207–218. https://doi.org/10.1111/j.1467-9280.1992.tb00029.x

Seidler, R. D. (2010). Neural correlates of motor learning, transfer of learning, and learning to learn. *Exercise and Sport Sciences Reviews*, *38*(1), 3–9. https://doi.org/10.1097/JES.0b013e3181c5cce7

Seidler, R. D., & Noll, D. C. (2008). Neuroanatomical correlates of motor acquisition and motor transfer. *Journal of Neurophysiology*, *99*(4), 1836–1845. https://doi.org/10.1152/jn.01187.2007

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: an integrative review. *Perspectives on Psychological Science*, *10*(2), 176–199. https://doi.org/10.1177/1745691615569000

St. John, M., Smallman, H. S., Manes, D. I., Feher, B. A., & Morrison, J. G. (2005). Heuristic automation for decluttering tactical displays. *Human Factors*, *47*(3), 509–525. https://doi.org/10.1518/001872005774860014

Stefanidis, D., Scerbo, M. W., Montero, P. N., Acker, C. E., & Smith, W. D. (2012). Simulator Training to Automaticity Leads to Improved Skill Transfer Compared With Traditional Proficiency-Based Training A Randomized Controlled Trial. *Annals of Surgery*, *255*(1), 30–37. https://doi.org/10.1097/SLA.0b013e318220ef31

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Szulewski, A., Howes, D., van Merriënboer, J. J. G., & Sweller, J. (2021). From Theory to Practice: The Application of Cognitive Load Theory to the Practice of Medicine. *Academic Medicine*, *96*(1), 24. https://doi.org/10.1097/ACM.0000000000003524

Teague, R. C., Gittelman, S. S., & Park, O. (1994). A Review of the Literature on Part-Task and Whole-Task Training and Context Dependency. Army Research Inst For The Behavioral And Social Sciences, Alexandria, VA. https://apps.dtic.mil/sti/citations/ADA285954

Thacker, I., Seyranian, V., Madva, A., Duong, N. T., & Beardsley, P. (2022). Social Connectedness in Physical Isolation: Online Teaching Practices That Support Under-Represented Undergraduate Students' Feelings of Belonging and Engagement in STEM. *Education Sciences*, *12*. https://eric.ed.gov/?id=EJ1339554

Thomas, I., Nicol, L., Regan, L., Cleland, J., Maliepaard, D., Clark, L., Walker, K., & Duncan, J. (2015). Driven to distraction: A prospective controlled study of a simulated ward round experience to improve patient safety teaching for medical students. *BMJ Quality & Safety*, *24*(2), Article 2. https://doi.org/10.1136/bmjqs-2014-003272

Vamadevan, A., Konge, L., & Bjerrum, F. (2024). Variable practice is superior to self-directed training for laparoscopic simulator training: A randomized trial. *Surgical Endoscopy*. https://doi.org/10.1007/s00464-024-10688-z

van Merriënboer, J. J. G., de Croock, M. B. M., & Jelsma, O. (1997). The transfer paradox: effects of contextual interference on retention and transfer performance of a complex cognitive skill. *Perceptual and Motor Skills*, *84*(3), 784–786. https://doi.org/10.2466/pms.1997.84.3.784

van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: recent developments and future directions. *Educational Psychology Review*, *17*(2), 147–177. https://doi.org/10.1007/s10648-005-3951-0

Wickens, C. D., Hutchins, S., Carolan, T., & Cumming, J. (2013). Effectiveness of part-task training and increasing-difficulty training strategies: a meta-analysis approach. *Human Factors*, *55*(2), 461–470. https://doi.org/10.1177/0018720812451994

Wightman, D. C., & Lintern, G. (1985). Part-task training for tracking and manual control. *Human Factors*, *27*(3), 267–283. https://doi.org/10.1177/001872088502700304

Wolfe, J. M. (2014). Approaches to visual search: Feature integration theory and guided search. *The Oxford Handbook of Attention*, *11*, 35–44.

Yao, L., DeVincenzi, A., Pereira, A., & Ishii, H. (2013). FocalSpace: Multimodal activity tracking, synthetic blur and adaptive presentation for video conferencing. *Proceedings of the 1st Symposium on Spatial User Interaction*, 73–76. https://doi.org/10.1145/2491367.2491377

Young, J. Q., Wachter, R. M., Cate, O. ten, O'Sullivan, P. S., & Irby, D. M. (2016). Advancing the next generation of handover research and practice with cognitive load theory. *BMJ Quality & Safety*, *25*(2), 66–70. https://doi.org/10.1136/bmjqs-2015-004181

Zepeda, C. D., Martin, R. S., & Butler, A. C. (2020). Motivational Strategies to Engage Learners in Desirable Difficulties. *Journal of Applied Research in Memory and Cognition*, *9*(4), Article 4. https://doi.org/10.1016/j.jarmac.2020.08.007

APPENDICES

Appendix A

List of Steps in Acquisition Videos

Acquisition Steps
1. Assemble Patient Circuit
1.1 Install disposable humidifier by placing it to the left of ventilator on top of the humidifier base/heater
1.2. Plug in humidifier heater
1.3 Replace heating element on humidifier heater
Vent flashes "air temperature okay"
1.4 Set up inspiratory circuit (no model, this is a subheading)
1.4.1.Disconnect short green tube from expiratory inlet
1.4.2 Connect other end of short green tube to either port on the disp. humidifier
1.4.3. Connect end of blue inspiratory tube to other port on disp. humidifier
1.4.4 Connect opposite end of white expiratory tube to the expiratory inlet
1.4.5. Connect end of white expiratory tube to yellow breathing filter
1.4.6. Disconnect yellow breathing filter from white expiratory tube
1.4.7 Connect green breathing filter to white expiratory tube
Vent flashes "breathing filter ok"
1.5. Connect the heater and inspiratory line
1.5.1. Connect yellow wire to yellow outlet on the heater
1.5.2. Connect other end of yellow wire to the blue inspiratory tube
1.5.3. Connect blue temperature probe to blue outlet on heater
1.5.4 Connect other end of blue temperature probe to blue inspiratory tube
1.5.5 Disconnect humidifier base/heater from AC Outlet
1.5.6. Connect orange humidifier valve to ventilator
1.5.7. Reconnect humidifier base/heater to AC Outlet
Vent flashes "humidifier temperature check, passed" for 10 seconds

Appendix B

List of Steps in Retention Test

Connect OxyPulse to Patient

Connect IVsupport to Patient

Give Syringe3 to Patient

Connect Dpads to Patient

Connect Vent to ElectricOut

Connect HumidBase to ElectricOut

Connect Hose1 from Vent to Port1

Connect Hose2 to Vent

Connect Hose2 to OxyTank

Connect Tube3 to ExIn

Connect Tube3 to InspOut

A/C Button

Press A/C Button (First Instance)

Press A/C Button (Second Instance)

Disconnect Vent from ElectricOut

Remove BattCov

Insert Batt

Replace BattCov

Connect Vent to ElectricOut

Press A/C Button (Third Instance)

Turn on O2valve

Remove Hose2 from OxyTank or O2valve

Connect Hose2 to Port1

Appendix C

List of Steps in Transfer Test

Connect pulse nitrometer to alien

Connect support fluid line to alien, give adrenaline shot

Connect emergency deamplifier to alien

Connect room filtration system to Type L outlet

Connect dehydrator base/air conditioner to Type L Outlet

Connect green valve from machine to wall port

Connect blue nitrogen hose from machine to orange oxygen tank

Connect the short blue pipe to expiratory inlet

Connect the short blue pipe to inspiratory outlet

Locate switch on side of machine

Turn machine on

Machine runs battery check

Turn machine off

Disconnect filtration system from Type L outlet

Remove Cell cover

Insert New Cell

Replace Cell cover

Reconnect ventilator to Type L outlet

Turn machine on

Air Quality Check

Turn on nitrogen tank

Remove blue nitrogen hose from orange oxygen tank

Connect blue nitrogen hose to wall port

Appendix D

Concept Retention Test

What is the purpose of giving a patient an IV line before a sedatives shot?

- ◯ In case there's an adverse reaction to the sedatives shot
- ◯ The sedative needs to react with the IV fluids to be effective
- ◯ The IV makes sure the patient is hydrated before the sedative shot
- ◯ The sedatives act as a backup to the IV in an emergency

What is the purpose of connecting the short green tube from the inspiratory outlet to the expiratory inlet?

- ◯ Air needs to be cooled before entering the patient
- ◯ The air needs to be purified before entering the patient
- ◯ The ventilator needs to be able to recycle air
- ◯ The short green tube has a backup air filter

We want to ensure that O2 levels don't drop during setup. What happens during the task to prevent this?

- ◯ Connecting the pulse oximeter
- ◯ Turning on the oxygen tank
- ◯ Passing the ventilator pre-check
- ◯ Connecting the red oxygen hose to the wall port

Why is the battery located in a separate compartment, in the front of the ventilator?

- ◯ For easy access in case of a power outage
- ◯ To seperate it from the pure oxygen in case of a short
- ◯ Because the air ports need to be located in the back for easy wall port access
- ◯ So the battery can be used in another ventilator, if needed

Why is the red hose connected to the oxygen tank, then connected to the wall port?

- ◯ Because the ventilator needs 100% oxygen to pass pre-check
- ◯ Because the oxygen tank is empty
- ◯ Because the oxygen tank has a leak
- ◯ Because the wall port delivers air more quickly

Why is the humidifier base/heater plugged in seperately?

- ○ Because heating the ventilator air requires more power
- ○ Because the water vapor could negatively impact the ventilator's main power supply
- ○ Because the humidifier base/heater could overheat and cause the ventilator to shut down
- ○ Because having both on one power cord may overload the outlets

What will the short green hose eventually be connected to?

- ○ The disposable humidifier
- ○ The green oxygen tank
- ○ The patient's breathing tube
- ○ The wall port

What is the correct order of steps for the automatic ventilator pre-check?

- ○ Battery check, oxygen check, air filter check
- ○ Air filter check, oxygen check, battery check
- ○ Oxygen check, battery check, air filter check
- ○ Battery check, air filter check, oxygen check

What two pieces are needed to check the air quality going to the patient?

- ○ Blue temperature probe and CO2 monitor
- ○ Yellow heater wire and blue temperature probe
- ○ CO2 monitor and humidity probe
- ○ Humidity probe and blue temperature probe

Why is the short green tube connected before turning on the ventilator?

- ○ Because the ventilator will start cycling air before being connected to the patient
- ○ Because the ventilator will cause an alarm without it
- ○ Because the pre-check will fail and restarting takes a long time
- ○ Because the ventilator won't turn on without it

Appendix E

Concept Transfer Test

What is the purpose of giving a patient a support fluid line before an adrenaline shot?

- ○ In case there's an adverse reaction to the adrenaline shot
- ○ The adrenaline needs to react with the support fluids to be effective
- ○ The support fluid line makes sure the patient is hydrated before the adrenaline shot
- ○ The adrenaline acts as a backup to the support fluid line in an emergency

What is the purpose of connecting the short blue pipe from the inspiratory outlet to the expiratory inlet?

- ○ Air needs to be cooled before entering the patient
- ○ The air needs to be purified before entering the patient
- ○ The room filtration system needs to be able to recycle air
- ○ The short blue pipe has a backup air filter

We want to ensure that nitrogen levels don't drop during setup. What happens during the task to prevent this?

- ○ Connecting the pulse nitrometer
- ○ Turning on the nitrogen tank
- ○ Passing the room filtration pre-check
- ○ Connecting the blue nitrogen hose to the wall port

Why is the room filtration system's cell located in a separate compartment, in the front?

- ○ For easy access in case of a power outage
- ○ To seperate it from the pure nitrogen in case of a short
- ○ Because the air ports need to be located in the back for easy wall port access
- ○ So the cell can be used in another filtration system, if needed

Why is the blue nitrogen hose connected to the nitrogen tank, then connected to the wall port?

- ○ Because the room filtration system needs 100% nitrogen to pass pre-check
- ○ Because the nitrogen tank is empty
- ○ Because the nitrogen tank has a leak
- ○ Because the wall port delivers air more quickly

Why is the dehydrator base/air conditioner plugged in separately?

- ◯ Because cooling the room filtration system air requires more power
- ◯ Because the dry air could negatively impact the room filtration system's main power supply
- ◯ Because the dehydrator base/air conditioner could overheat and cause the room filtration system to shut down
- ◯ Because having both on one power cord may overload the outlets

What will the short blue pipe eventually be connected to?

- ◯ The disposable dehydrator
- ◯ The orange nitrogen tank
- ◯ The patient's breathing tube
- ◯ The wall port

What is the correct order of steps for the automatic ventilator pre-check?

- ◯ Cell check, nitrogen check, air filter check
- ◯ Air filter check, nitrogen check, cell check
- ◯ Nitrogen check, cell check, air filter check
- ◯ Cell check, air filter check, nitrogen check

What two pieces are needed to check the air quality going to the patient?

- ◯ Yellow temperature regulator and N2O monitor
- ◯ Orange cooling wire and yellow temperature regulator
- ◯ N2O monitor and aridity probe
- ◯ Aridity probe and yellow temperature regulator

Why is the short blue pipe connected before turning on the room filtration system?

- ◯ Because the room filtration system will start cycling air before being connected to the patient
- ◯ Because the room filtration system will cause an alarm without it
- ◯ Because the pre-check will fail and restarting takes a long time
- ◯ Because the room filtration system won't turn on without it

Appendix F

ADHD Self Report Scale (Kessler et al., 2005)

1. How often do you have trouble wrapping up the final details of a project, once the challenging parts have been done?
    a. Never, Rarely, Sometimes, Often, Very Often
2. How often do you have difficulty getting things in order when you have to do a task that requires organization?
    a. Never, Rarely, Sometimes, Often, Very Often
3. How often do you have problems remembering appointments or obligations?
    a. Never, Rarely, Sometimes, Often, Very Often
4. When you have a task that requires a lot of thought, how often do you avoid or delay getting started?
    a. Never, Rarely, Sometimes, Often, Very Often
5. How often do you fidget or squirm with your hands or feet when you have to sit down for a long time?
    a. Never, Rarely, Sometimes, Often, Very Often
6. How often do you feel overly active and compelled to do things, like you were driven by a motor?
    a. Never, Rarely, Sometimes, Often, Very Often
7. How often do you make careless mistakes when you have to work on a boring or difficult project?
    a. Never, Rarely, Sometimes, Often, Very Often
8. How often do you have difficulty keeping your attention when you are doing boring or repetitive work?
    a. Never, Rarely, Sometimes, Often, Very Often
9. How often do you have difficulty concentrating on what people say to you, even when they are speaking to you directly?
    a. Never, Rarely, Sometimes, Often, Very Often
10. How often do you misplace or have difficulty finding things at home or at work?
    a. Never, Rarely, Sometimes, Often, Very Often
11. How often are you distracted by activity or noise around you?
    a. Never, Rarely, Sometimes, Often, Very Often
12. How often do you leave your seat in meetings or other situations in which you are expected to remain seated?
    a. Never, Rarely, Sometimes, Often, Very Often
13. How often do you feel restless or fidgety?
    a. Never, Rarely, Sometimes, Often, Very Often
14. How often do you have difficulty unwinding and relaxing when you have time to yourself?
    a. Never, Rarely, Sometimes, Often, Very Often
15. How often do you find yourself talking too much when you are in social situations?
    a. Never, Rarely, Sometimes, Often, Very Often
16. When you're in a conversation, how often do you find yourself finishing the sentences of the people you are talking to, before they can finish them themselves?

    a. Never, Rarely, Sometimes, Often, Very Often
17. How often do you have difficulty waiting your turn in situations when turn taking is required?
    a. Never, Rarely, Sometimes, Often, Very Often
18. How often do you interrupt others when they are busy?
    a. Never, Rarely, Sometimes, Often, Very Often

Appendix G

Attentional Control Scale (Derryberry & Reed, 2002)

1. It's very hard for me to concentrate on a difficult task when there are noises around.
    a. Almost Never, Sometimes, Often, Always
2. When I need to concentrate and solve a problem, I have trouble focusing my attention.
    a. Almost Never, Sometimes, Often, Always
3. When I am working hard on something, I still get distracted by events around me.
    a. Almost Never, Sometimes, Often, Always
4. My concentration is good even if there is music in the room around me.
    a. Almost Never, Sometimes, Often, Always
5. When concentrating, I can focus my attention so that I become unaware of what's going on in the room around me.
    a. Almost Never, Sometimes, Often, Always
6. When I am reading or studying, I am easily distracted if there are people talking in the same room.
    a. Almost Never, Sometimes, Often, Always
7. When trying to focus my attention on something, I have difficulty blocking out distracting thoughts.
    a. Almost Never, Sometimes, Often, Always
8. I have a hard time concentrating when I'm excited about something.
    a. Almost Never, Sometimes, Often, Always
9. When concentrating I ignore feelings of hunger or thirst.
    a. Almost Never, Sometimes, Often, Always
10. I can quickly switch from one task to another.
    a. Almost Never, Sometimes, Often, Always
11. It takes me a while to get really involved in a new task.
    a. Almost Never, Sometimes, Often, Always
12. It is difficult for me to coordinate my attention between the listening and writing required when taking notes during lectures.
    a. Almost Never, Sometimes, Often, Always
13. I can become interested in a new topic very quickly when I need to.
    a. Almost Never, Sometimes, Often, Always
14. It is easy for me to read or write while I'm also talking on the phone.
    a. Almost Never, Sometimes, Often, Always
15. I have trouble carrying on two conversations at once.
    a. Almost Never, Sometimes, Often, Always
16. I have a hard time coming up with new ideas quickly.
    a. Almost Never, Sometimes, Often, Always
17. After being interrupted or distracted, I can easily shift my attention back to what I was doing before.
    a. Almost Never, Sometimes, Often, Always
18. When a distracting thought comes to mind, it is easy for me to shift my attention away from it.
    a. Almost Never, Sometimes, Often, Always

19. It is easy for me to alternate between two different tasks.
    a. Almost Never, Sometimes, Often, Always
20. It is hard for me to break from one way of thinking about something and look at it from another point of view.
    a. Almost Never, Sometimes, Often, Always

Appendix H

Medical Knowledge Survey

Have you ever participated in a ventilation procedure?

- ◯ Yes
- ◯ No

Have you ever witnessed a ventilation procedure?

- ◯ Yes
- ◯ No

Rate your knowledge of medical ventilation procedures.

- ◯ Excellent
- ◯ Good
- ◯ Average
- ◯ Poor
- ◯ No knowledge

Which of the following ventilators are you familiar with?

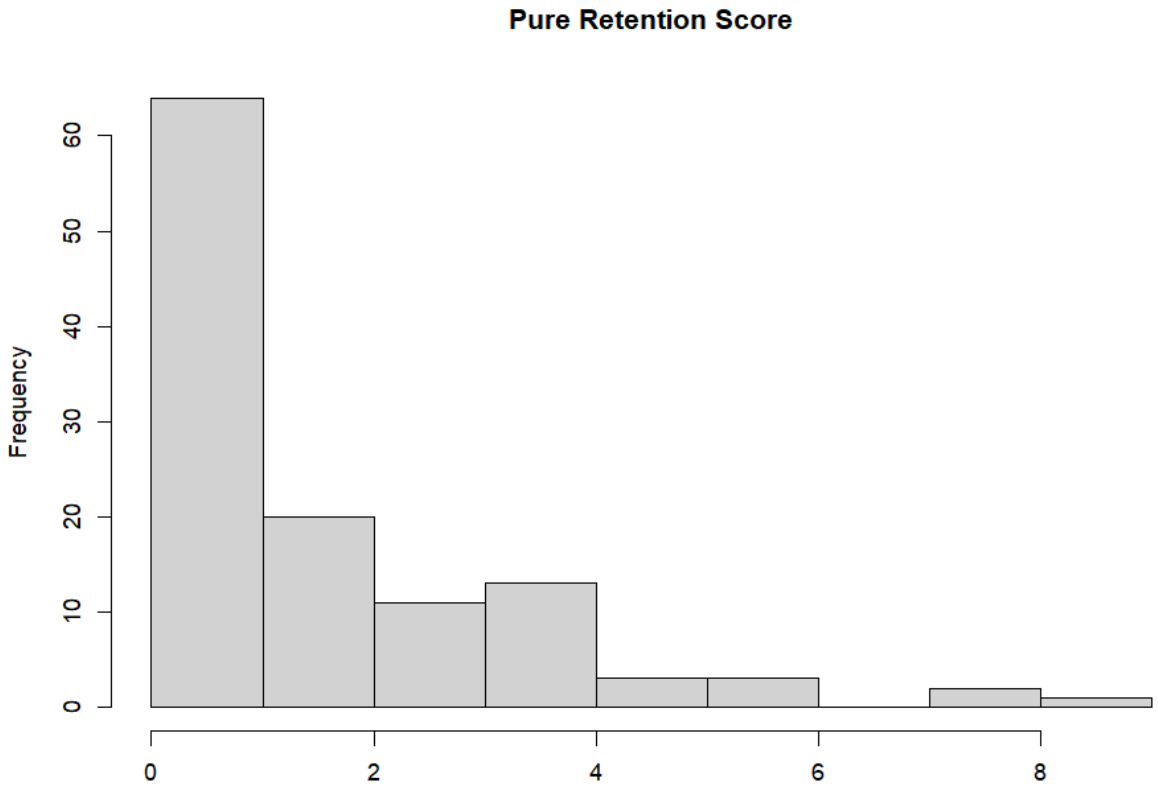
- ☐ Negative pressure
- ☐ Intensive care
- ☐ Positive pressure
- ☐ Transport
- ☐ None of the above

What is a medical procedure that might require ventilation? If you are unsure type, "I don't know."

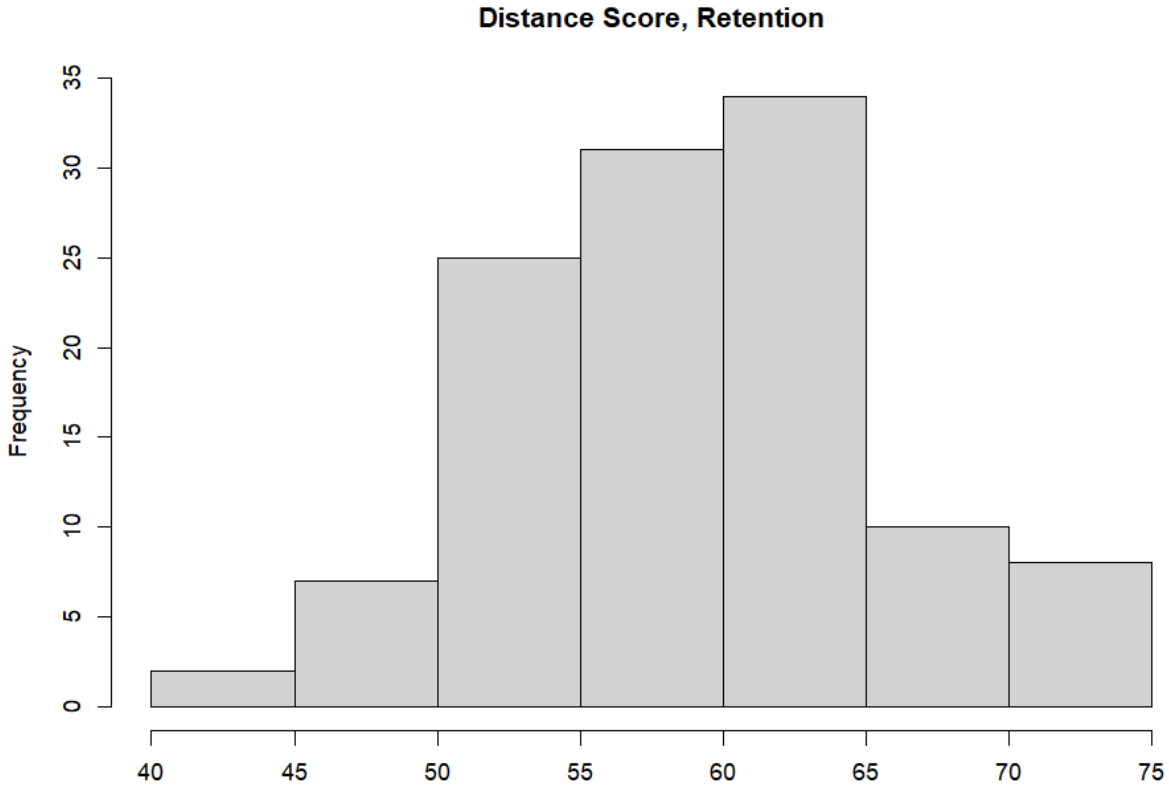
How many years have you worked in the medical field? If none, put 0.

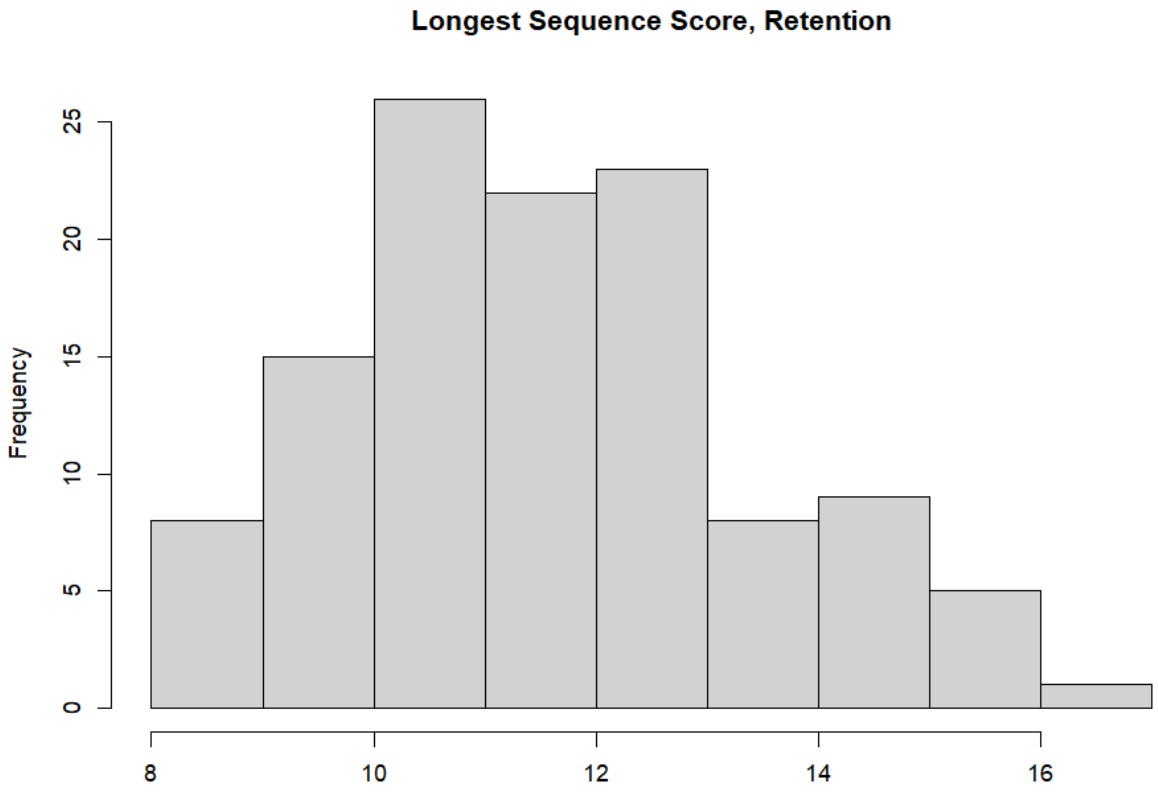
Appendix I

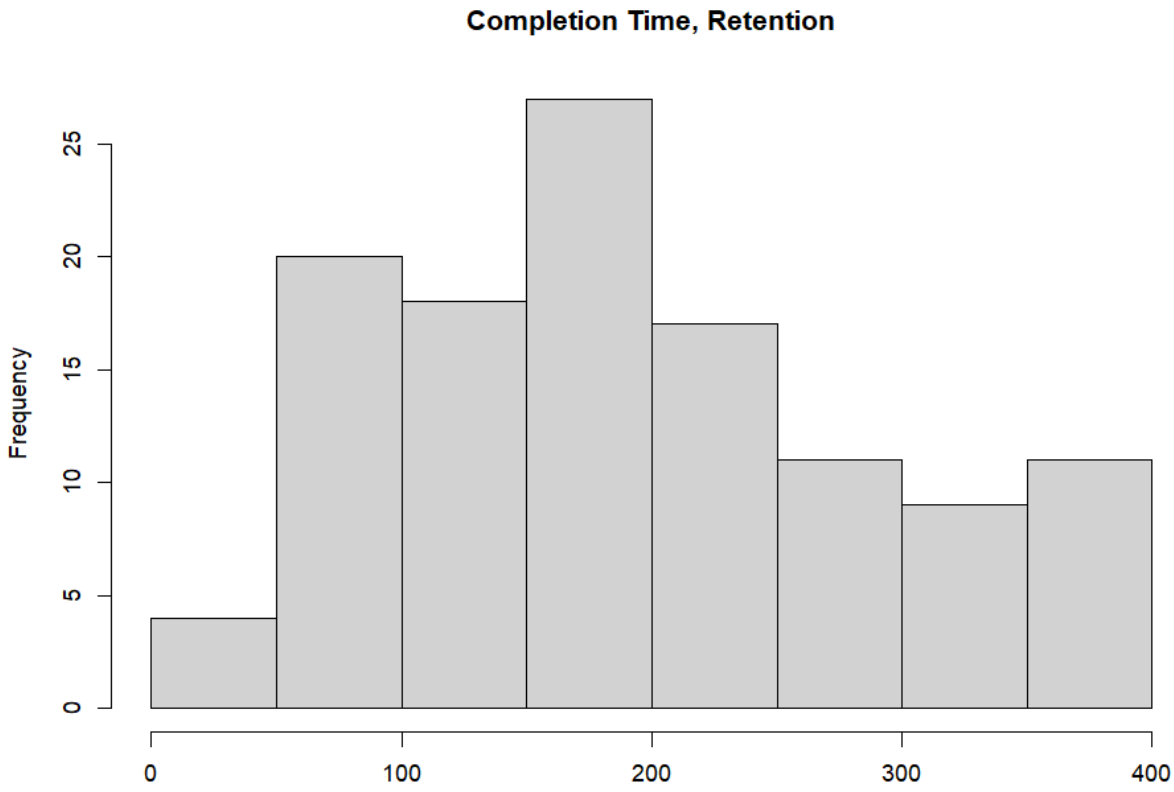
Histograms of Retention Outcome Variables




Pure Retention Score, Log Transformation
Frequency
0
5
10
15
20
25
30
35
0.0
0.5
1.0
1.5
2.0

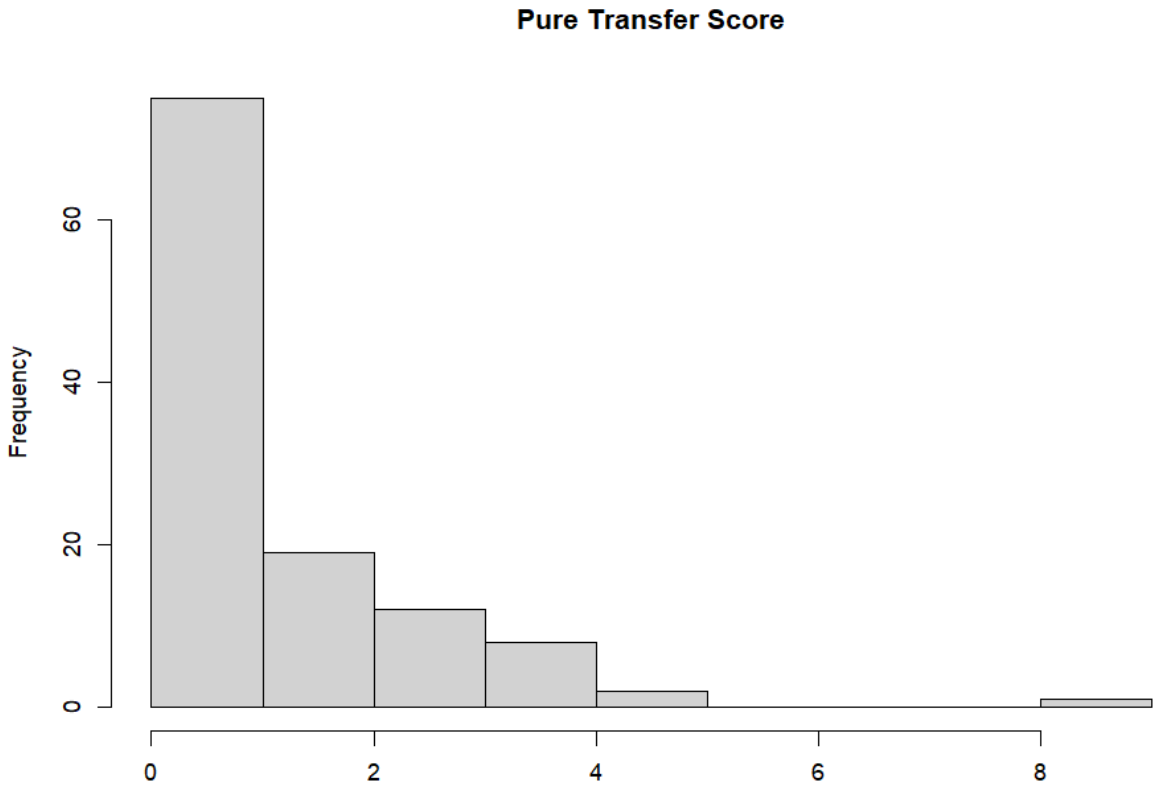

Distance Score, Retention
Frequency
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75

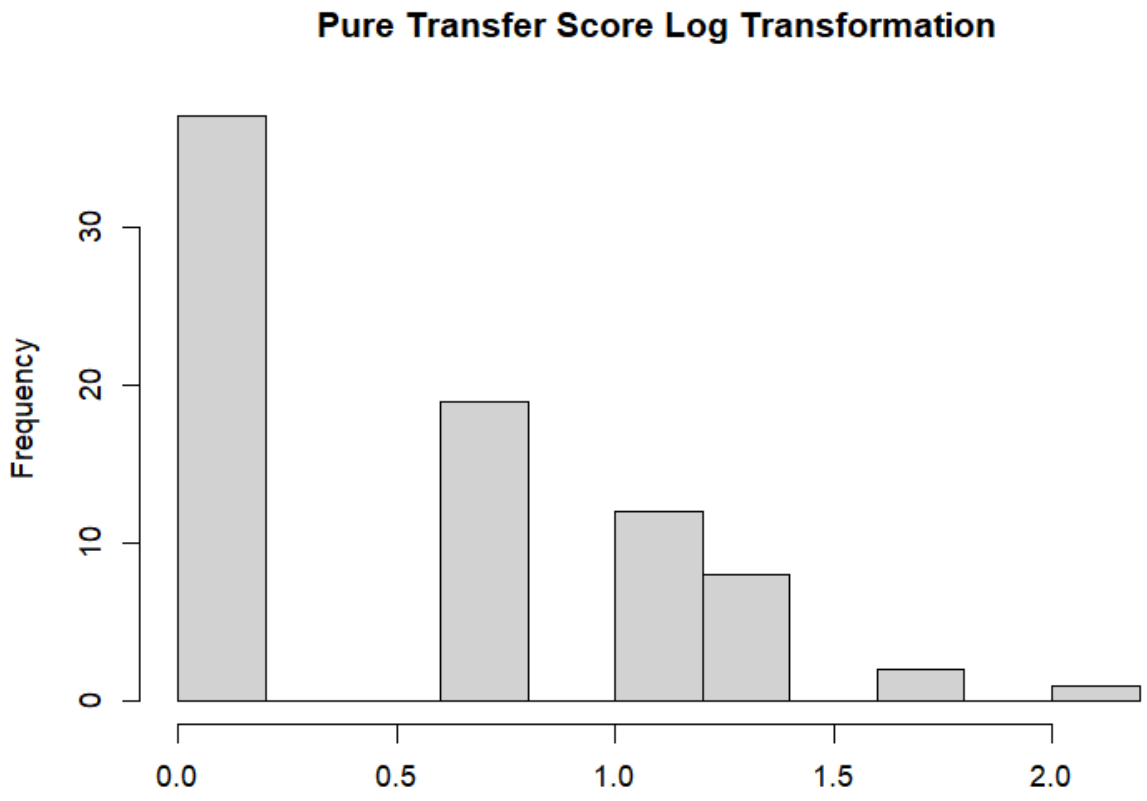

Longest Sequence Score, Retention
Frequency
0
5
10
15
20
25
8
10
12
14
16

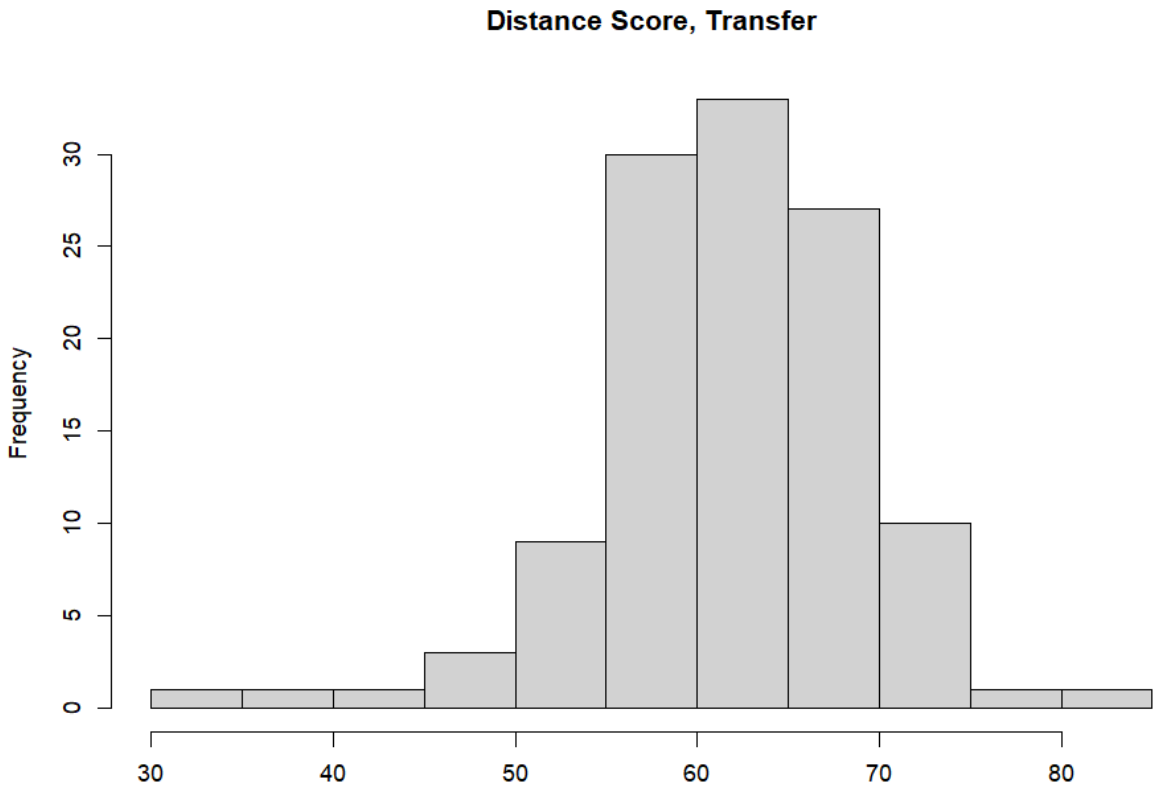

Completion Time, Retention
Frequency
0
5
10
15
20
25
0
100
200
300
400

Appendix J

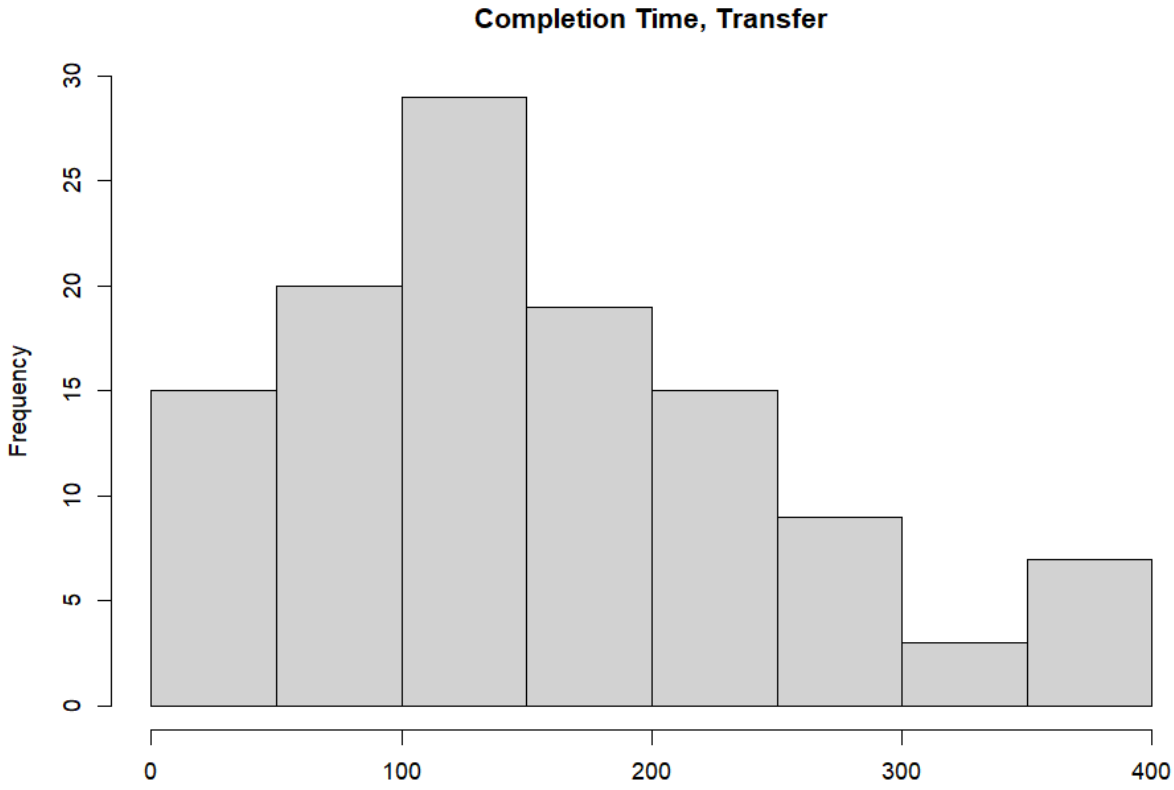
Histograms of Transfer Outcome Variables




Pure Transfer Score Log Transformation
Frequency
0
10
20
30
0.0
0.5
1.0
1.5
2.0


Distance Score, Transfer
Frequency
0
5
10
15
20
25
30
30
40
50
60
70
80


Longest Sequence Score, Transfer
Frequency
0
5
10
15
20
25
30
8
10
12
14
16


Completion Time, Transfer
Frequency
0
5
10
15
20
25
30
0
100
200
300
400

Appendix K

Description of Distance and Longest Sequence Measures

Retention Score

Retention score was the strictest scoring method used, where participants received 1 point for each step that was in the correct place. For example, if step 1 was listed first in a participant's response, they received one point. If step 1 was listed anywhere else, they received zero points. After scoring the retention test and observing the low performance across all groups, we considered alternative scoring methods that rewarded participants for partially correct answers. Distance score can be best described as rewarding participants for being close having steps in the right place, and longest sequence score can be described as rewarding participants for remembering chunks of the steps.

Distance Score

Distance score allowed for participants to gain partial credit if they were near where they should be (i.e., if step 5 is in slot 5, the most credit is given. Less would be given if it is in slot 7, and even less if it were in slot 23). The numerical distance from the correct step added to participants' score, which meant that a higher score indicated worse performance. For example, if step 2 was placed in slot 20, 19 would be added to the score). It was possible to get a score higher than 23 (the total number of steps), if there were multiple steps that were placed far away from the correct step.

Longest Sequence Score

Longest sequence allowed for credit if steps were in the correct order, even if they were not in the correct slots. For example, if steps 5, 6, 7, and 8 followed each other directly, credit

was given even if that string of steps were in slots 12, 13, 14, and 15. The highest possible score would be 23, for having all the steps in the correct order.