



DFRWS 2020 USA — Proceedings of the Twentieth Annual DFRWS USA

Exploring the Learning Efficacy of Digital Forensics Concepts and Bagging & Tagging of Digital Devices in Immersive Virtual Reality

Courtney Hassenfeldt ^{a, b, c, e, *}, Jillian Jacques ^{a, b, d}, Ibrahim Baggili ^{a, b}^a Connecticut Institute of Technology at University of New Haven, USA^b University of New Haven Cyber Forensics Research and Education Group (UNHcFREG), Samuel S. Bergami Jr. Cybersecurity Center, USA^c Cybersecurity & Networks Graduate Student, USA^d Computer Science Undergraduate Student, USA^e MITRE Corporation, USA

ARTICLE INFO

Article history:

Keywords:

Virtual reality
Cyber forensics
Digital forensics
Education
Gamification

ABSTRACT

This work presents the first account of evaluating learning inside a VR experience created to teach Digital Forensics (DF) concepts, and a hands-on laboratory exercise in Bagging & Tagging a crime scene with digital devices. First, we designed and developed an immersive VR experience which included a lecture and a lab. Next, we tested it with ($n = 57$) participants in a controlled experiment where they were randomly assigned to a VR group or a physical group. Both groups were subjected to the same lecture and lab, but one was in VR and the other was in the *real world*. We collected pre- and post-test results to assess the participants' knowledge in DF concepts learned. Our experimental results indicated no significant differences in scores between the immersive VR group and the physical group. However, our results showed faster completion times in VR by the participants, which hints at VR being more time efficient, as virtual environments can be spun programmatically with little downtime.

© 2020 The Author(s). Published by Elsevier Ltd on behalf of DFRWS. All rights reserved. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The world is changing and evolving to become more technology focused; from cars that drive themselves, to smart houses and pocket sized computing devices. All these changes have been integrated into our daily lives. Learning is no different. In 2015, \$13 billion were spent on hardware technology for education purposes (Munster et al., 2015). Technological integration into learning started with laptops and online databases for research, and now schools have moved to fully online programs. Yet, many argue that online programs cannot replace the immersive experience one receives in a physical classroom.

This is where Immersive Virtual Reality (VR) comes into play. At the time of writing, remote learning had become critical to the survival of academic institutions due the 2019–2020 Corona Virus Disease (COVID-19) pandemic. This pandemic made it clear that remote instruction is imperative. While remote instruction may be

feasible in lecture style settings, teaching concepts that rely on physical space, such as Bagging & Tagging digital devices from a crime scene, is difficult to conduct without hands-on, tactile, experiential learning. Immersive VR may help in such instances.

While it is expected that the education sector will result in 700,000 VR headset units being sold in 2021 (Consulting, 2017), with prices decreasing over time (Munster et al., 2015), questions remain unanswered about the efficacy of employing VR in Digital Forensics (DF) and cybersecurity education. We aimed to improve the state of the art and decrease the knowledge gap in this domain. As such, our work presents the following contributions:

- To the best of our knowledge, we present the first openly available VR DF education game via the Immersive VR Education ENGAGE platform¹.
- We conduct the first experimental study to explore DF learning in immersive VR versus a physical (real world) learning environment.

The paper is organized as follows. Section 2 describes

* Corresponding author. University of New Haven Cyber Forensics Research and Education Group (UNHcFREG), Samuel S. Bergami Jr. Cybersecurity Center, USA.

E-mail addresses: chasst1@unh.newhaven.edu (C. Hassenfeldt), jjacq2@unh.newhaven.edu (J. Jacques), ibaggili@newhaven.edu (I. Baggili).

¹ <https://engagevr.io/>

background information and related work. Section 3 outlines the methodology used to deploy the study and Section 4 discusses the results obtained from the completion of the study. Next, Section 5 details the limitations and Section 6 reviews key findings and discussion. Finally, Section 7 contains our conclusions and future work.

2. Background and related work

2.1. Digital forensics education

Shinder and Cross (2002) and Whitcomb (2002) were among the first that identified the necessity for DF science. At the time, the development was primarily driven by agencies and vendors. Rogers and Seigfried (2004) then conducted a needs analysis survey showing that education, training, certification, and research funding in DF were among the top reported challenges. These results also coincide with recent work (Harichandran et al., 2016; Luciano et al., 2018).

Gottschalk et al. (2005) examined computer forensic efforts of colleges and universities in the United States. They concluded that DF is a growing field. At the time, few traditional education programs were available, with a notable researched program at Champlain College (Kessler and Schirling, 2006). In most programs at the time, students learned investigation techniques, and did not gain experience from working on real-world cases as part of their curriculum McGuire and Murff (2006); Conklin (2006).

Kessler (2007), however, was the first to describe some of the course design aspects of teaching computer forensics in an online environment with particular attention to the issue of hands-on learning in an online environment. In his work, he highlighted the importance of hands-on tasks.

As research in DF education progressed, Taylor et al. (2007) performed an assessment of DF programs. The authors concluded that many schools at the time only provided one or two general courses of study, or offered concentrations. They again stressed the importance of hands-on exercises. Further work by Nance et al. (2010) presented an education agenda for DF and stressed the importance of educational materials that consist of practical experience, case studies, assessment items, and scenarios.

Lang et al. (2014) proposed a DF curriculum which emphasized the importance of balancing training and education. They stated that professional education and certification led to development of training-based courses that teach DF as a stepwise laboratory procedure, and have neglected to educate students in the theoretical foundations of what they are learning.

Most relevant to our work were two recently published papers that created simulated learning environments. The first discussed designing a Virtual Crime Scene Simulator that included the ability to interact with, and perform live triaging of commonly-found digital devices (Conway et al., 2015). This experience was gamified, but was not explored in immersive VR. The other work, by Karabiyik et al. (2019), focused on designing a VR educational experience framework for digital forensic first responders. Their developed framework focused on creating a simple student scoring mechanism, and was in VR. However, they simply discussed the design of their VR framework and conducted no formal evaluation of their approach to test its efficacy in learning.

A thorough review of the literature in DF education shows that although research has been conducted in the domain, there has been little to no work performed on the creation, learning evaluation, and dissemination of an immersive simulated, situated, DF VR learning environment that mimics the real world.

2.2. Education and assessment theory

Educational theory is constantly impacted by new technology, and learning in immersive environments is grounded in past seminal work. To cover all the foundational research in education and assessment theory is beyond the scope of this paper, but we touch on some of the most relevant topics briefly in this section.

In pedagogy, immersive authentic learning aligns with constructivist and situated learning theory (Herrington et al., 2010). Immersion places a student in a simulated, or real-world physical and social context while guiding, scaffolding (instructional support to students during learning), and facilitating participatory metacognition (Palincsar, 1998). These processes include authentic inquiry, active observation, peer coaching, and reciprocal teaching (Squire, 2010).

Of importance in educational theory is authentic inquiry in simulated learning environments (Windschitl, 2003) and situated learning theory. Situated learning theory involves learning activities that are embedded in and inseparable from participation in physical, social, or cultural settings (Brown et al., 1989; Lave et al., 1991). Educational literature also suggests that optimal learning opportunities should support scaffolding and responsive feedback as well as active reflections as students complete the learning activity (Simon et al., 2006).

Past work by Dede (2009) demonstrated that immersive learning environments in virtual worlds can enhance and augment education by enabling situated learning, providing students with a complex social learning space and time for reflective practice in an optimal environment (Alessi and Trollip, 2001).

2.3. VR education work in other domains

Recent work has illustrated the efficacy of the use of VR education for students to practice dissections in the medical domain (Kiourexidou et al., 2019). VR has also been used to educate patients on the risks of certain diseases, with results showing that they retained knowledge directly after, 1 week after, and 1 year after the VR experience (Balsam et al., 2019). Lastly, of relevance to our work is that VR has been employed in some information security classrooms where researchers found that students were more motivated to learn using VR than other educational mediums (Ma, 2018).

We note that when designing our VR experience presented in Section 3, educational theory and educational best practices (such as authentic learning, immersion, automatic feedback, and simulated environments) were employed to ensure that the constructed learning environment would be optimized for learning.

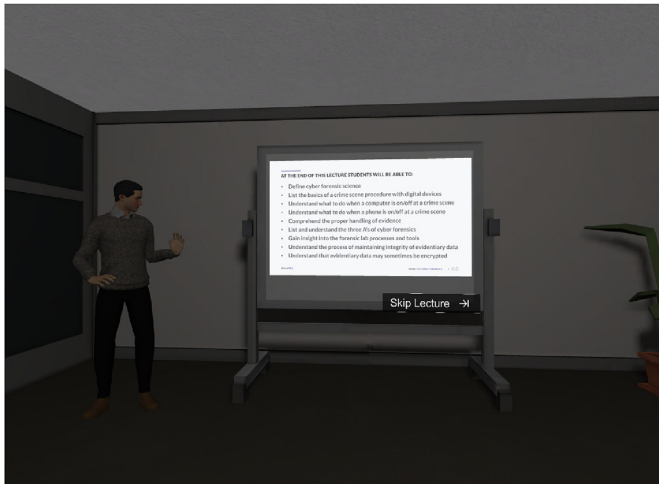
3. Methodology

In this section, Table 1 outlines the apparatus utilized, the experience creation, how participants were selected, the setup of the physical and virtual labs, and the pre- and post-tests. The following overarching methodology was employed to conduct this study:

1. The VR experience (game) was designed by our team, developed by Immersive VR Education, and placed on the ENGAGE platform.
2. IRB paperwork was submitted and approved.
3. We distributed to students the experiment opportunity via various sources (E-mail, university instructors, and our psychology department).
4. The experimental study was conducted by randomly assigning participants to the VR group or physical group. Participants completed a multiple choice test related to DF topics before

Table 1
Apparatus.

Hardware/software	Version	Use
HTC Vive	N/A	To run VR experience on
Computer	Windows 10	To run VR system
SteamVR	1.9	To run ENGAGE
ENGAGE	1.1.9	To run VR experience
Mac Desktop	Windows 10 (dual-boot)	For students to take pre/post tests and physical group to watch lecture on
Qualtrics	N/A	To create and distribute pre/post tests
SPSS	25	To perform statistical analysis

**Fig. 1.** Screenshot: VR view - Professor giving lecture.**Fig. 2.** Screenshot: VR view - CEO giving task to investigator.

entering the learning environment, as well as after completion of the learning environment. At the conclusion of each experiment, the participants were debriefed.

Our hypotheses were as follows:

- H1.** There will be a statistically significant difference between pre- and post-test scores.
- H2.** There will be no statistically significant difference in pre- and post-test scores between the VR group and the physical group.
- H3.** There will be a statistically significant difference between the average completion time of participants in the VR lab and in the physical lab.

3.1. VR experience creation

The first step taken was to create the Bagging & Tagging VR experience, based on a lab exercise typically conducted in PI Baggili's Introduction to Digital Forensics course. This was carefully thought-out and designed to specifications outlined by the research team. The physical lab was then created to mirror a replica of the VR experience. A game resembling the laboratory environment, called *Office Agent*, was designed by our team. The requirements were then shared with Immersive VR Education, and the experience was developed and placed freely on the ENGAGE platform. The ENGAGE platform is a VR education platform that showcases educational lectures and Immersive VR experiences. The goal was to create a situated, simulated, learning environment that followed educational best practices. To complete the VR experience, the following steps were taken:

- The educational experience was designed to include a lecture and lab by graduate student Daniel Walnycky and PI Baggili, and implemented by Immersive VR Education.
- A lecture audio was studio recorded and then edited into a video with its respective PowerPoint slides. This lecture was then published on YouTube.
- The lecture then served as the medium for the VR experience, however, an avatar for the professor was created as well (Fig. 1).
- The laboratory experience was then designed to include a briefing by a CEO for solving a case (Fig. 2), and an on-scene investigation for an evidence Bag & Tag exercise (Figs. 3 and 4).

The lecture covered basic DF topics and lasted 20 min and 22 s². It focused on types of evidence, methods and tools for imaging/extraction/acquisition, tools and software for forensic analysis, and hash values. After the lecture, students would have an understanding of evidence handling at a crime scene, including the steps necessary to maintain evidence integrity and a chain of custody. Participants would also gain an understanding of the importance of the three As of DF: Acquire, Authenticate, and Analyze.

Next, the gamified laboratory experience implemented by ENGAGE placed participants in the shoes of an investigator and provided a scenario meant to resemble that of the real world. The experience begins with an introduction from a CEO explaining that he believes there is an information leak in his company, and that stolen blueprints are still somewhere in the office. Participants are tasked with finding the stolen blueprints, as well as obtaining any pertinent evidence in a forensically sound manner, while the office

² The lecture can be found here: <https://www.youtube.com/watch?v=gjtk-j-5yJA&t=1s>



Fig. 3. Screenshot: VR view - left side of scene.



Fig. 4. Screenshot: VR view - right side of scene.

employees are out on a company retreat.

To demonstrate the concept of forensic soundness and the importance of following forensic procedures, participants must complete certain tasks that are provided when collecting specific pieces of evidence. For example, Figs. 7 and 8 show how a player is to correctly organize the steps for collecting a flash drive. When submitted in the correct order, a green mark is provided, otherwise a red mark is provided. Players are able to try a variety of orders until the correct one is found. This VR experience provided the participant a chance to explore the concepts learned from the lecture.

3.2. IRB approval process

For any work to be completed with human participants, Institutional Review Board (IRB) approval was needed. For this project, an application was completed and submitted to the University of New Haven IRB under Protocol Number: 2019-081. The approval process took approximately one month. The dates allotted to conduct the experiment were November 6, 2019 to November 1, 2020. The data that was submitted to the review board included: a completed application, lecture material, pre- and post-tests, test

answers, a consent form to be signed by participants, ideal sample sizes, the site used to randomly assign groups, distribution information, debriefing statements, and certifications. The PI and all involved researchers completed the appropriate Collaborative Institutional Training Initiative (CITI) training.

3.3. Participant recruitment

Gaining feedback from participants with various backgrounds and fields of study allowed for more generalizable results on the efficacy of our constructed VR educational experience. People learn differently and come from different backgrounds. Those wanting to learn DF may not always come from computing or DF backgrounds, therefore our study was intentionally inclusive of all backgrounds.

To recruit participants with varying backgrounds, three major approaches were taken to circulate the study opportunity:

- We distributed the study opportunity via an e-mail to all students and faculty.
- We arranged a time frame of 3 h to sit at a table at the entrance of a popular cafe on campus to engage with those who walked by.
- We disseminated and displayed flyers across campus in different buildings as a way to reach all majors including biology, criminal justice, arts and sciences, engineering, and psychology.

All of the aforementioned approaches advertised that participants would receive a \$5 Dunkin' Donuts gift card. Participants were then placed into random groups (VR & Real World) using the site randomlists.com.

3.4. Virtual Lab Setup

Studies have demonstrated that room temperature has an effect on a participant's ability to accomplish a task. For this reason, a neutral temperature of 72 °F was selected to keep the environment the same among all tests conducted (Energy Air, 2018).

Once the temperature was set, the HTC Vive VR headset and controllers, and the computer used for participants to complete pre- and post-tests were disinfected in front of participants. This was to maintain a clean and stable environment across all tests, as well as to assure participants that they were using clean tools. Participants were then instructed to pick a random sheet of paper from 80 squares that were face down and contained varying participation numbers, and to keep the number to themselves and not share with the Principle Investigator (PI) or Co-PI. This participation ID was recorded when completing the pre- and post-tests online to keep track of the participant's responses.

Once an ID was obtained by the participant, they completed the pre-test using the given computer in the lab. This computer was separate from the computer used for the VR setup and had the test already prepared on the screen for the participant. After completing the pre-test, the lecture was viewed on YouTube. The lab was then completed in VR, followed by the post-test. For analysis purposes, the post-test consisted of the same questions posed in the pre-test, including the participant ID. After the post-test was completed, the participant was debriefed, and their ID was shredded to ensure that the ID was never tied to the individual.

3.5. Physical lab setup

The physical lab setup was similar to that of the VR lab, utilizing the same room and setting the room temperature to 72 °F. Additionally, the keyboard used by the participants to complete the pre-



Fig. 5. Physical lab setup.

and post-tests was also wiped down with disinfectant wipes. Materials from the physical crime scene, however, were not wiped down after the initial setup as participants were required to wear gloves throughout their tasks.

To start, participants watched the same YouTube video as those in the VR lab group. To ensure accurate results for comparison, the physical lab mirrored the environment in the VR lab game. This was achieved with the addition of a desk, desktop computer, chair, filing cabinet, CD, paper, chair, and other items seen in the game, including a picture frame on the desk (Fig. 5). This simulation attempt can be seen in Figs. 3 and 4, which are of the VR game setting, and Fig. 5, which is of the physical lab setup. Additionally, questions asked in the VR game were also mirrored in the physical lab by utilizing magnetic cards on a white board. Similar to the VR game, when participants were asked questions in the physical lab, the magnetic cards were used to provide answers.

4. Results

4.1. Demographics

There was a total of ($n = 57$) participants. The majority of the participants were white (71.9%) females (64.9%) ranging from 18 to 24 years old (94.7%). Almost 90% of participants had a high school diploma (50.91%) or some college, but no degree (38.6%). Most respondents were in the criminal justice & forensic sciences program (56.1%) with a field of study of criminal justice (45.6%) and an estimated GPA between 3.1 and 3.5 (45.6%) or between 3.6 and 4.0 (28.1%).

In terms of employment status, 38 out of the 57 respondents (66.7%) listed their status as being a student while 17 participants indicated being a student and working (29.8%) and only 2 of the participants defined themselves as not being a student and working (3.5%). When asked to indicate their field of occupation or the field they would like to enter, 52.6% selected a field related to criminal justice or law enforcement (33.3% Criminal Justice or Forensic Sciences, 8.8% Government, Public Administration or Military, 10.5% Law, Public Safety, Corrections, Security).

The participants' demographics are exemplified in Table 2.

4.2. Interest in DF and cybersecurity

Interest in cybersecurity was surveyed utilizing four Likert scale

questions. The first question measured familiarity with cyber/digital forensics concepts. 63.2% of participants fell in the range of strongly disagree to somewhat disagree, 15.8% neither agreed nor disagreed, and 21.1% fell between somewhat agree and strongly agree. The next question focused on the interest in cyber/digital forensics as a career path to which 47.4% fell on the side of disagree, 29.8% neither agreed nor disagreed, and 22.8% agreed. The next question centered on salary to see if it played a role in participants interest in cyber/digital forensics careers. For this, it was found that 79% of participants either somewhat agreed, agreed, or strongly agreed while 5.3% either disagreed or strongly disagreed and 15.7% neither agreed nor disagreed.

Lastly, participants were asked if they were interested in having a career that makes the world a safer place. 54 out of 57 participants somewhat agreed, agreed, or strongly agreed compared to the 3 out of 57 participants that somewhat disagreed (1 participant) or neither agreed nor disagreed (2 participants). While a majority of our participants were females (37 out of 57), when asked this question all 37 fell in the range of somewhat agree to strongly agree, while the only people who answered neither agree nor disagree, or somewhat disagree were males. This coincides with the ideas proposed by Woodcock et al. (2013) that "women are typically more oriented toward people and men more oriented toward things".

4.3. Pre-test

To test the efficacy of the lecture and lab provided to the participants, a pre- and post-test were given; this was used to measure the information learned. There were a total of 15 multiple choice questions with four potential answers involving introductory digital forensics concepts. These questions were answered by all 57 participants. The test scores were calculated out of 15 ($M = 7.26$, $SD = 1.92$). There was a minimum score of 3 and a maximum score of 12, with a range of 9.

The first question asked was "Cyber forensics can be defined as;" which 61.4% of participants answered correctly with, "the scientific examination and analysis of data held on, or retrieved from, computer storage media". The next question, "what is a digital crime scene?" was answered correctly by 75.4% of participants. Almost 60% of participants correctly answered the question "regardless of whether a digital device is on or off, when entering a crime scene it's important to first:" with "take pictures of the crime scene". Participants were then asked, "what is an important extra step in the cyber forensics process when a computer is found turned ON at a crime scene?". For this question the correct answer was "acquire the RAM," which 47.4% of participants chose.

In relation to the question "What is an important extra step in the cyber forensics process when a smartphone is found turned ON at a crime scene?" only two of the four answer choices were selected by participants. The two choices were "remove any SIM cards or SD cards" and "put device in airplane mode and connect it to a portable charge". Of the participants, 21.1% answered this correctly with "put device in airplane mode and connect it to a portable charge". Question number six asked participants "in any case that you work on you should always approach it as if ____" to which two answer choices were most commonly chosen. Those two choices were "the evidence will reach the court of law", selected by 45.6%, and "time is limited", selected by 42.1%. The correct answer here was "the evidence will reach the court of law".

When asked, "what is of the highest importance during evidence collection?" only 10 out of 57 participants answered correctly with "forensic soundness of collected evidence", while the majority (33 out of 57 participants) answered "finding all potential evidence". A majority of participants answered the question "what

Table 2
Demographics.

Gender	
Female	64.9%
Male	33.3%
Gender Fluid	0.02%
What is the highest level of school you have completed or the highest degree you have received?	
High school graduate (high school diploma or equivalent including GED)	50.9%
Some college but no degree	38.6%
Associate degree in college (2-year)	3.5%
Bachelor's degree in college (4-year)	5.3%
Master's degree	1.8%
Race	
White	71.9%
Black or African American	10.5%
American Indian or Alaska Native	3.5%
Asian	5.3%
Native Hawaiian or Pacific Islander	1.8%
Other	7.0%
In your most recent educational experience, what program were your under?	
Arts and Sciences	26.3%
Business	5.3%
Criminal Justice & Forensic Sciences	56.1%
Engineering	3.5%
Fine Arts	1.8%
Other	7.0%
In your most recent educational experience, what was your major/field of study?	
Business	1.8%
Communication	3.5%
Computer Science	3.5%
Criminal Justice	45.6%
Cyber Systems	1.8%
Environmental Science	1.8%
Finance	1.8%
Forensic Science	3.5%
Hospitality & Tourism Management	1.8%
Music	1.8%
Music Industry	3.5%
Music and Sound Recording	7.0%
Nutrition & Dietetics	1.8%
Psychology	1.8%
Undeclared/Exploratory	5.3%
Other	14.0%
In your most recent educational experience, what was your estimated GPA?	
Between 0.0 and 1.0	1.8%
Between 2.1 and 3.0	23.2%
Between 3.1 and 3.5	46.4%
Between 3.6 and 4.0	28.6%
Which statement best describes your current employment status?	
Student	66.7%
Student + Working	29.8%
Not a Student + Working	3.5%

do the “three A’s” in cyber forensics stand for?” correctly with “Acquire, Authenticate, and Analyze” (54 out of 57 participants). The next question “In forensics the term Image means a forensically sound copy of a hard drive. What do you use to ensure the integrity of a hard drive disk image?” had the correct response from 51 out of 57 participants which was “both A and B” (a hardware write blocker and computational hash).

In terms of the question “What is hashing?”, the answers were somewhat evenly distributed among three of the four answer choices. Those choices being “the encryption of a disk image to prevent unauthorized access” at 24.6% of participants, “the bypassing of encryption to gain access to a suspect disk image” at 33.3% of participants, and the correct answer of “the transformation of a file into a fixed-length alpha-numeric value that acts as a unique identifier to the original file” being slightly higher at 35.1% of participants. The following question asked, “what is one of main challenges of analyzing hard drives?”, to which 68.4% of participants answered correctly with “encryption”. It was then asked, “list tools widely used in the cyber forensics industry for hard drive disk analysis:”. The majority answered incorrectly with “Genymotion,

Eclipse, and Virtual Box” (33.3%) while the correct answer, “Autopsy, Access Data, and EnCase”, received 28.1% of participant’s answers.

The question “list the four types of small scale digital device forensic acquisition:” with the correct answer of “Manual, Logical, File System, Physical” was correctly answered by 29.8% of responses. The last pre-test question was “In what situation would you want to acquire a small scale digital device physical image?”. 6 out of 57 participants answered correctly with “When needing to find deleted/escalated privilege files”, while the leading answer choice amongst participants was “Both A and B” (when needing to find deleted/escalated privilege files and when needed to analyze the file system of a device).

4.4. Post-test

Participants were asked the same 15 questions in their post-test for comparison purposes. The average number of correct answers was almost 10 ($M = 9.96$ $SD = 1.79$). The minimum score received was a 7 and the maximum score was a 14. A paired T-Test was run to

compare the scores from the pre- and post-tests. Results indicated that there was a mean increase of 2.7 (SD = 2.02) and that there was a significant statistical difference between the pre- and post-test scores ($p = 3.0647 \times 10^{-14}$, $t = 10.111$, $DF = 56$, $t(43) = 4.85$). The paired samples correlation results show a positive correlation value of .412.

A Univariate Analysis of Variance was then run to see if there was a correlation between which group a participant was placed in and the difference in score between their pre- and post-tests. This resulted in no significant statistical difference ($F = 0.526$, $DF = 1$, $p = 0.471$). Gender by group was also tested to see if this affected test scores. We found that there was no statistically significant difference ($F = 0.265$, $DF = 4$, $p = 0.899$) for this either.

When asked again, "Cyber forensics can be defined as: ", the number of participants who answered correctly rose 15.8% with 77.2% now getting the question correct. The next question, "what is a digital crime scene?" was answered correctly by 78.9% of participants, which was 3.5% greater than previously. All but two participants answered the question "regardless of whether a digital device is on or off, when entering a crime scene it's important to first:" correctly on the post-test. The question, "what is an important extra step in the cyber forensics process when a computer is found turned ON at a crime scene?" was correctly answered by 34 participants which is seven more than previously.

For the question "What is an important extra step in the cyber forensics process when a smartphone is found turned ON at a crime scene?" the same two answer choices were selected among all participants. However, in the post-test the majority of participants chose the correct answer (82.5%). Thirteen more participants correctly answered "in any case that you work on you should always approach it as if ____" than previously, bringing the total to 39 out of 57.

A question that stuck out in terms of results was "what is of the highest importance during evidence collection?". Even after the lecture and lab, a majority of participants answered incorrectly. Only 15 out of 57 participants answered correctly, which is 5 more than on the pre-test.

When asked "what do the three As in cyber forensics stand for?" all participants answered correctly. This is the only question that all participants answered correctly in the post-test. For the next question, "In forensics the term Image means a forensically sound copy of a hard drive. What do you use to ensure the integrity of a hard drive disk image?" 41 out of 57 participants answered correctly, which is 10 less than on the pre-test. However, it should be noted that while less got this correct than before, only two participants chose answer C. This is of significance because the correct answer was "Both A and B". While participants might have gotten the answer incorrect, a majority chose answer choices that were at least somewhat correct.

The question "what is hashing?" previously had 35.1% of participants answering correctly, but after the lab and lecture that rose to 59.6%. Next, when asked, "what is one of main challenges of analyzing hard drives?" 82.5% answered correctly, 14.1% more than on the pre-test. When asked for the second time "list tools widely used in the cyber forensics industry for hard drive disk analysis:", the amount of participants that correctly answered almost tripled (increasing to 40 out of 57 participants).

Twenty-six more participants correctly answered the question "list the four types of small scale digital device forensic acquisition:" during the post-test than the pre-test, bringing the total up to 43 out of 57. The last question in the post-test was "In what situation would you want to acquire a small scale digital device physical image?". Three less participants answered this correctly than before, with the leading answer still being "Both A and B" (when needing to find deleted/escalated privilege files and when needed

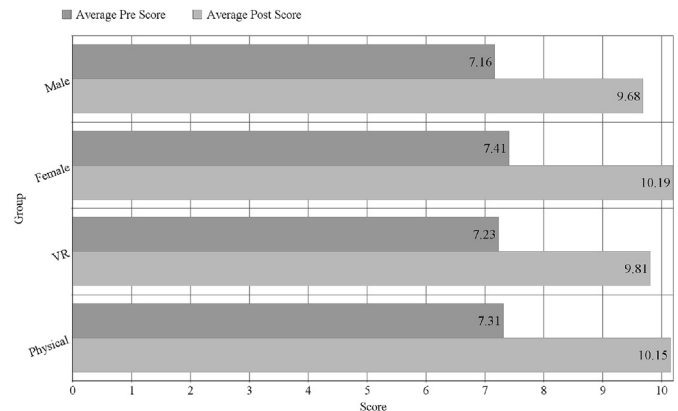


Fig. 6. Bar graph showing the Pre- and Post- test score averages by gender and lab group.

to analyze the file system of a device).

4.5. Gender

There was no statistically significant difference between the male and female group's test scores. This was concluded after multiple ANOVA tests were run to check the difference between pre- and post-test scores of both groups. For the females pre-test there was a slightly higher average than the male group ($M = 7.41$, $SD = 1.83$). Though males scored an average of 0.25 lower ($M = 7.16$, $SD = 2.03$), this was not significantly different ($p = 0.211$). The post-test scores also had a higher average for females ($M = 10.19$, $SD = 1.86$) than males ($M = 9.68$, $SD = 1.53$) with a ($p = 0.151$).

4.6. VR lab group versus physical lab group

Participants were randomly selected to complete either the physical lab or the VR lab as explained in section 3. To see if there was a statistically significant difference between groups, the results were compared. The results between the VR and physical groups, as well as genders, can be seen in the graph in Fig. 6.

The average pre-test scores for the VR lab group and the physical lab group were close, with the VR lab score being ($M = 7.23$, $SD = 2.32$) and the physical lab score being ($M = 7.31$, $SD = 1.35$). There were 26 participants placed in the physical lab group and 31 participants placed in the VR lab group. An ANOVA test was run, and resulted in there being no statistically significant difference ($DF = 1$, $F = 0.025$, and $p = 0.874$).

The average time it took participants to complete the VR lab was 10 min and 33 s. The average time for participants to complete the physical lab was 14 min and 18 s, roughly 4 min more than participants in the VR lab. There are a couple factors that could have played a part in this discrepancy. First, the VR lab could not go longer than 13 min due to limitations discussed in Section 5. A second factor that could have contributed to the increased time was the ordering section. As discussed in section 3, the physical lab mirrored the VR lab as closely as possible so there was an ordering component, as seen in Figs. 7 and 8. Unlike in the VR lab, the researchers had to check the ordering and mark the answers right or wrong for the participant. This contributed to the time recorded.

When taking a look at the post-test averages, it was found that the VR lab group had a similar average score ($M = 9.81$, $SD = 1.62$) as the physical lab group, who came in just a touch higher in their averages ($M = 10.15$). An ANOVA test indicated that the difference between the two groups was not statistically significant ($DF = 1$, $F = 0.527$, and $p = 0.471$).

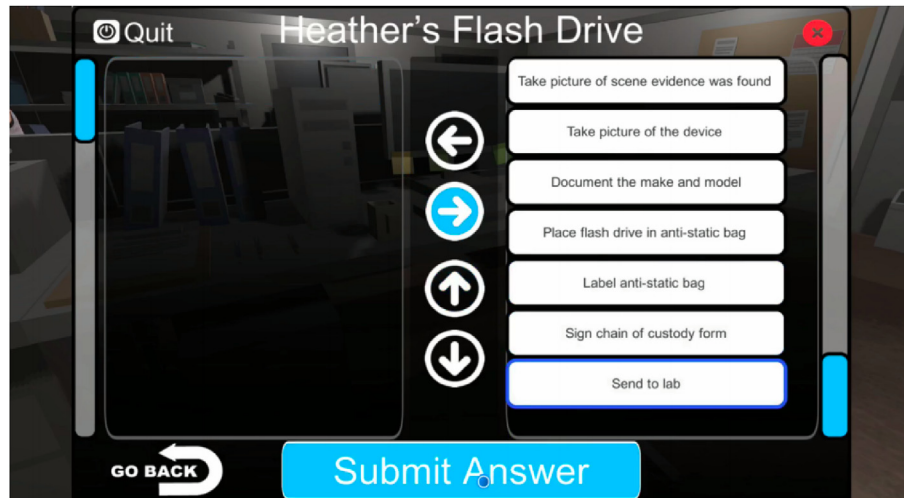


Fig. 7. Screenshot: VR view - flash drive ordering.

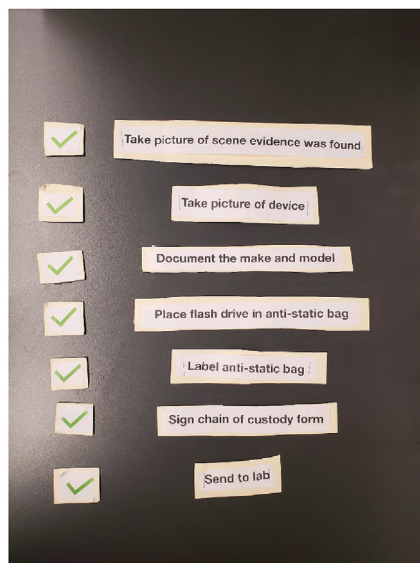


Fig. 8. Physical lab flash drive ordering.

4.7. Overall lab experience question

The participants were provided with the opportunity to discuss what they liked most and least about the experience. Many participants reported enjoying the VR experience and the realistic feel of it. Some feedback from participants included: "The person lecturing at the beginning was really interesting to experience", "I liked how everything looked real as I was walking around", and "It was my first VR experience so it was something new. It felt like I was in a different room because everything was so realistic. I also learned a lot about cyber forensics that I had no knowledge of before." The group who experienced the physical lab also provided feedback about having enjoyed the lab and lecture, including that "(t)he hands on experience emulat(ed) the feeling of being in a crime scene."

Participants reported that they least liked the length of the lecture and would have preferred to get to the lab faster. However, some noted that while they thought this, they understood the lecture was a necessity. Another issue reported was that the VR was

slightly blurry which made it hard to read the lecture or that the headset was uncomfortable, for instance: "The lecture was a bit blurry and I somewhat focused on that instead of the lecture", "The goggles were quite heavy, and the screen made me a bit dizzy.", and "I did not like the discomfort of the headset".

During the lab experience there were not many technical difficulties, but some participants did report a couple things. The main technical difficulty reported for the VR group was having trouble getting used to VR and how it works with responses such as, "at first i had some trouble with the virtual controls (pointing and clicking), but i was able to improve with them." The only technical issue that occurred for the physical lab group was that during the exercise they locked the filing cabinet while searching for evidence.

There was then an allotted time for providing suggestions for improvement. Many responses from those who participated in the VR experience would have liked time to practice using VR before completing the lab. A few participants who completed the physical lab would have preferred it was completed in a group.

While the lack of exposure to VR is expected to be a confounding variable, this was part of the intended experimental setup as a way to test if participants still learned without having previous VR experience. The results indicated that participants, whether having exposure to VR, or not, increased their test score between pre- and post-tests.

Likert scale questions were then asked to gauge the participant's feelings towards the lecture questions and experience. First, they were asked to rate the statement, "the quiz after the lecture was challenging." The results ranged from agree to strongly disagree with 28 out of 56 respondents evenly split between somewhat disagree or disagree. 71.9% of participants agreed or somewhat agreed that this study required a great deal of focus. Almost half of participants agreed that they learned a lot from this experience, while over half agreed that the instructor of the lecture was knowledgeable of the material. Twenty-four of the 57 participants neither agreed nor disagreed with the statement, "this experience has encouraged me to consider a career in cyber forensics".

4.8. VR experience question

These next questions asked were just for the ($n = 31$) participants that completed the VR lab, which for many was their first VR experience (71%). The experience was completed in multiple ways: 25.8% completed it sitting down, 12.9% standing, 32.3% a mix of

sitting and standing, and 29% sat for the lecture and stood for the lab. Only 16.1% of participants fully wore glasses and 3.2% partially wore glasses during the lab.

Over 90% of the VR participants agreed to some degree that it felt as if they were in the virtual room (Strongly Agree = 45.2%, Agree = 22.6%, Somewhat Agree = 25.8%). While the next question asking to rate the statement, “the virtual teacher looked and acted like a real person”, received more of a variation of results, the majority fell on the side of agreeing to some degree (22 out of 31 participants). The highest rating for the statement “I was engaged and focused on the lecture” was agree with 35.5%. Many agreed to some degree that they would encourage their friends to participate in a VR lecture (20 out of 31 participants). When asked to rate the question “the VR experience made me dizzy, lightheaded, or nauseous”, 29% agreed to some extent, 12.9% neither agreed nor disagreed, and 58.1% disagreed to some extent. There was an even spread of results for the question asking to rate the statement “virtual reality lectures should be integrated into college courses” (Strongly Agree = 16.1%, Agree = 16.1%, Somewhat Agree = 19.4%, Neither Agree nor Disagree = 19.4%, Somewhat Disagree = 16.1%, Disagree = 6.5%, Strongly Disagree = 6.5%).

5. Limitations

There were a few limitations that should be noted. First, there was little variation in age groups, 94.7% were aged 18–24 and the remaining 5.3% were in the 25–39 age group. Participants in different age groups may have provided different perspectives on VR. For example, older participants may have experienced more dizziness in VR, and may have not been able to complete the study. Nonetheless, the study was aimed at college-age groups as they would most likely be the users of VR in educational settings.

Another limitation was that the VR experience timed out after roughly 13 min due to the VR experience being a trial version. With that said, the VR participants still showed an increase in their post-test scores.

One more limitation is that the experience was designed to focus on physical activities, such as Bagging & Tagging a crime scene - the results may vary if the VR experience had students work on a computer system, investigating files, or writing code. Another variable of our study that may be seen as a limitation is the higher percentage of female participants. However, our results indicated no statistically significant differences of scores between male and female participants.

6. Discussion

Our work showed that given a specific environment and laboratory exercise in Bagging & Tagging a DF crime scene, there was no significant difference between the scores of students that learned in VR versus physical space. What is important to note is that our results also showed that the completion time for the laboratory exercise was lower in VR, which is expected, as VR environments can be pragmatically replicated.

Companies are now moving towards using VR systems to teach cybersecurity awareness, and our results may be generalizable to that domain, however, further testing would be needed as the results may vary between VR experiences.

From 2003 to 2012, the FBI saw over double the amount of cases that needed some form of digital examinations (Me, 2014), and the workforce in DF is expected to grow by 28 percent from 2016 to 2026 (BLS, 2019). Our work may inspire a novel approach for attracting new talent to the area of DF since it has been previously shown that students were motivated to learn more using VR in an information security class (Ma, 2018).

Provided a technology like VR, simulated, situated learning environments may be useful in educating the future workforce in this domain. Training may become more efficient and can be provided to law enforcement practitioners who are seeing the increase in digital evidence.

In Fall 2016 alone, there were 6.3 million students that had enrolled in at least online classes (Friedman, 2018). While many universities are moving towards online programs, students may have a more difficult time grasping material when it is difficult to perform physical labs that can only be done in person. VR could be the solution to that, especially as headsets and systems become more affordable and widespread.

7. Conclusion & future work

Our results allowed us to accept *H1*, *H2*, and *H3*. For *H1*, we found that, overall, there was a statistically significant increase in test scores from participant's pre- and post-test scores. In terms of *H2*, it was found that there was no significant difference in test scores among those who participated in the VR lab versus the physical lab, supporting the idea that both methods are viable practices. Lastly, with *H3* we did find a statistically significant difference between the time it took for participants to complete the VR lab versus physical lab with the VR lab requiring less time.

These results may indicate that employing VR to teach DF (at least within the context of lectures and Bagging & Tagging laboratory exercise) is an effective approach. This is beneficial due to the widespread growing popularity of VR headsets which means teaching of DF on a larger scale may be possible.

Since results showed that there was no statistically significant difference between the pre- and post-test scores of the participants that completed the VR experience versus the physical experience, we contend that this reinforces the idea that for a tactile task, such as Bagging & Tagging, a VR experience can be used to teach such skills. Purchasing a VR headset could be a cheaper, more feasible alternative method of teaching such topics, while yielding similar learner comprehension.

Future work should focus on extending the VR labs to include more training and different immersive experiences. This should include lectures of more in depth material, as well as more extensive labs. We also hope to test in the future the VR environment with law enforcement practitioners and gain feedback on how experiences may more closely resemble the real world. Other ideas for future work would be to test this type of learning on participants while teaching more complex and technical computing-based tasks such as DF tool usage, and programming exercises.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1748950. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We would also like to acknowledge all the participants that took part in the study and Daniel Walnycky for his help designing the VR experience and for Immersive VR Education for the implementation of the VR experience. We would also like to thank Dr. W. Ian O'Byrne, Assistant Professor of Literacy Education, College of Charleston, for guiding us on educational theory and assessment. Lastly, we would like to thank Laura Sanchez, Cyber Security Engineer, MITRE, for proof-reading our work. Additional Disclaimer: Approved for Public Release: Distribution Unlimited, 20-1659. Author Hassenfeldt's affiliation with The MITRE Corporation is provided for identification purposes only, and is not intended to convey or imply MITRE's

concurrence with, or support the positions or viewpoints expressed by the author.

References

- Alessi, S.M., Trollip, S.R., 2001. *Multimedia for Learning: Methods and Development*. Allyn & Bacon.
- Balsam, P., Borodzicz, S., Malesa, K., Puchta, D., Tymirńska, A., Ozierański, K., Koitowski, Ł., Peller, M., Grabowski, M., Filipiak, K.J., et al., 2019. Oculus study: virtual reality-based education in daily clinical practice. *Cardiol. J.* 26, 260–264.
- BLS, 2019. *Information Security Analysts : Occupational Outlook Handbook*.
- Brown, J.S., Collins, A., Duguid, P., 1989. Situated cognition and the culture of learning. *Educ. Res.* 18, 32–42.
- Conklin, A., 2006. Cyber defense competitions and information security education: an active learning solution for a capstone course. In: *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, vol. 9. IEEE, 220b–220b.
- Consulting, F., 2017. *Futuresource Consulting Press: Vr in Education*.
- Conway, A., James, J.I., Gladyshev, P., 2015. Development and initial user evaluation of a virtual crime scene simulator including digital evidence. In: *International Conference on Digital Forensics and Cyber Crime*. Springer, pp. 16–26.
- Dede, C., 2009. Immersive interfaces for engagement and learning. *Science* 323, 66–69.
- Energy Air, I., 2018. Does Classroom Temperature Affect Learning?.
- Friedman, J., 2018. Study: More Students Are Enrolling in Online Courses.
- Gottschalk, L., Liu, J., Dathan, B., Fitzgerald, S., Stein, M., 2005. Computer forensics programs in higher education: a preliminary study. In: *Proceedings of the 36th SIGCSE Technical Symposium on Computer Science Education*, pp. 147–151.
- Harichandran, V.S., Breitering, F., Baggili, I., Marrington, A., 2016. A cyber forensics needs analysis survey: revisiting the domain's needs a decade later. *Comput. Secur.* 57, 1–13.
- Herrington, J., Reeves, T.C., Oliver, R., 2010. *A Guide to Authentic E-Learning*. Routledge, New York.
- Karabiyik, U., Mousas, C., Sirota, D., Iwai, T., Akdere, M., 2019. A virtual reality framework for training incident first responders and digital forensic investigators. In: *International Symposium on Visual Computing*. Springer, pp. 469–480.
- Kessler, G.C., 2007. Online education in computer and digital forensics: a case study. In: *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*. IEEE, 264a–264a.
- Kessler, G.C., Schirling, M.E., 2006. The design of an undergraduate degree program in computer & digital forensics. *J. Digital Forensics Secur. Law* 1, 3.
- Kiourexidou, M., Piagkou, M., Ndunju Njau, S., Natsis, K., Antonopoulos, N., Bamidis, P., 2019. Multimedia application. *Aristotle Biomed. J.* 1.
- Lang, A., Bashir, M., Campbell, R., DeStefano, L., 2014. Developing a new digital forensics curriculum. *Digit. Invest.* 11, S76–S84.
- Lave, J., Wenger, E., et al., 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge university press.
- Luciano, L., Baggili, I., Topor, M., Casey, P., Breitering, F., 2018. Digital forensics in the next five years. In: *Proceedings of the 13th International Conference on Availability, Reliability and Security*, pp. 1–14.
- Ma, W.-m., 2018. Impact of incorporating virtual reality into information security forensic teaching activities on learning motivation. *Int. J. Tourism Leisure Cult. Creative Fash. Des. (IJTLCCFD)*.
- McGuire, T.J., Murff, K.N., 2006. Issues in the development of a digital forensics curriculum. *J. Comput. Sci. Coll.* 22, 274–280.
- Me, G., 2014. Migration to governmental cloud digital forensics community: economics and methodology. In: *International Conference on Grid Economics and Business Models*. Springer, pp. 114–129.
- Munster, Gene, Jakel, Travis, Clinton, Doug, Murphy, Erinn, 2015. *Next Mega Tech Theme Is Virtual Reality*. PiperJaffray.
- Nance, K., Armstrong, H., Armstrong, C., 2010. Digital forensics: defining an education agenda. In: *2010 43rd Hawaii International Conference on System Sciences*. IEEE, pp. 1–10.
- Palincsar, A.S., 1998. Social constructivist perspectives on teaching and learning. *Annu. Rev. Psychol.* 49, 345–375.
- Rogers, M.K., Seigfried, K., 2004. The future of computer forensics: a needs analysis survey. *Comput. Secur.* 23, 12–16.
- Shinder, D.L., Cross, M., 2002. *Scene of the Cybercrime*. Elsevier.
- Simon, S., Erduran, S., Osborne, J., 2006. Learning to teach argumentation: research and development in the science classroom. *Int. J. Sci. Educ.* 28, 235–260.
- Squire, K., 2010. From information to experience: place-based augmented reality games as a model for learning in a globally networked society. *Teach. Coll. Rec.* 112, 2565–2602.
- Taylor, C., Endicott-Popovsky, B., Phillips, A., 2007. Forensics education: assessment and measures of excellence. In: *Second International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE'07)*. IEEE, pp. 155–165.
- Whitcomb, C.M., 2002. An historical perspective of digital evidence: a forensic scientist's view. *Int. J. Digital Evid.* 1, 7–15.
- Windschitl, M., 2003. Inquiry projects in science teacher education: what can investigative experiences reveal about teacher thinking and eventual classroom practice? *Sci. Educ.* 87, 112–143.
- Woodcock, A., Graziano, W.G., Branch, S.E., Habashi, M.M., Ngambeki, I., Evangelou, D., 2013. Person and thing orientations: psychological correlates and predictive utility. *Soc. Psychol. Pers. Sci.* 4, 116–123.