Exploring Effects of Interactivity on Learning with Interactive Storytelling in Immersive Virtual Reality

Lei Zhang
Human-Centered Design
Virginia Tech
Blacksburg, USA
leiz@vt.edu

Doug A. Bowman
Dept. of Computer Science
Virginia Tech
Blacksburg, USA
dbowman@vt.edu

Caroline N. Jones
Dept. of Biological Sciences
Virginia Tech
Blacksburg, USA
jonescn@vt.edu

***Abstract*—Immersive virtual reality (VR) holds great potential for learning, but it is unclear how VR experiences should be designed to maximize learning potential. In this study, we explored how the level of interactivity in an educational VR storytelling experience for immunology learning affects a user's learning gains. We created three versions of the VR experience with low (system automates as many actions as possible), medium (a combination of system automation and user-controlled actions), and high (as many user-controlled actions as possible) levels of interactivity. We hypothesized that too much or too little interactivity would result in smaller learning gains than a medium level of interactivity. Although data from pre and post-tests showed no significant difference in students' learning gains due to interactivity level, questionnaire and interview data suggest that interactivity in the experience significantly affects students' engagement in learning, attention, and focus on learning material. Participants also perceived that they could learn better and more effectively in a VR experience with a higher level of interactivity.**

Keywords—virtual reality, interactivity, learning, storytelling, instructional design, immunology, education

I. Introduction

Imagine yourself as a white blood cell within a human body, where you patrol in a blood vessel with other immune cells as your routine duty. You are alerted of a body infection and have to find a portal allowing you to migrate from the blood vessel into body tissues. There you use different weapons to fight off invading bacteria and control infection. This is the scenario of a recent educational VR storytelling experience we developed to teach college-level students specific immunology concepts [1]. Using immersive VR technology, we were able to visualize the concepts in the context of an interactive narrative and create an engaging learning experience for students.

The practice of using VR for education is not new. As early as the 1990s, VR projects like ScienceSpace [2] used immersive VR to teach students abstract concepts of mechanics, electromagnetism, and chemistry. For quite some time, researchers and educators have been exploring how to leverage some unique affordances of VR, such as interactivity, immersion, embodiment, multisensory stimuli, and the first-person point of view (POV) to enhance various learning experiences. Studies on those topics have suggested that VR is "an extremely promising tool for the enhancement of learning, education, and training" [3].

However, although studies have shown positive evidence that VR can promote learning [2], there are no solid guidelines in the literature on how to leverage VR affordances to design VR experiences for maximum learning effectiveness. In particular, few studies have addressed how the interactivity provided by VR experiences affects learning outcomes, or how much interactivity is needed for optimal learning effectiveness in a virtual environment (VE). In our VR application, we rely on interactivity and storytelling as the primary design elements to convey abstract learning concepts to learners. Thus, we are faced with a critical design choice: how much interactivity to provide to users for maximum learning.

It is not always true that providing more interactivity in instructional materials will lead to improved learning outcomes. Studies on multimedia interactivity and cognitive load suggest that too much learner control and choice may result in extraneous interactivity that hinders the learning process due to the limited capacity of working memory [4] [5]. Therefore, we hypothesize that there is a "sweet spot" in the interactivity design for educational VR experiences where an appropriate amount of interactivity in the experience is just enough and can lead to optimal learning effectiveness.

In this paper, we report on an experiment that investigated how the level of interactivity affects learning performance, especially in the context of interactive VR storytelling for complex science concept learning. The experiment used our immunology learning experience as a testbed, since it is a complete VR learning experience that can be modified to provide different levels of interactivity. We compared three versions of the experience with low, medium, and high interactivity levels.

Results from our experiment showed that there was no significant difference in students' objective learning gains from pre- and post-tests between VR experiences with different levels of interactivity. However, evidence from subjective questionnaires revealed that with an increased level of interactivity, students reported significantly more engagement, concentration, and efficiency in learning. They also believed that medium to high levels of interactivity in the experience helped them effectively understand immunology concepts.

Our primary contributions are:

- We provide empirical evidence that helps guide interaction design in storytelling/narrative-driven interactive VR experiences for learning.
- We demonstrate the impact of interactivity level on learning performance and user experience in VR.

- We provide guidelines to help educational VR designers make decisions on how much interactivity to be embedded in a VR experience for optimal learning.

II. Related Work

A. VR's Potential to Support Learning

Practices and studies of using VR to promote learning have been abundant in the fields of education and instructional design since the 1990s. Some unique characteristics of VR technology, such as 3D spatial representations, scientific and data visualization, multisensory user interaction, and immersion distinguish it from other interactive instructional media and make it an ideal tool to support learning of specific concepts. In a review of VR applications in education, Slater [3] summarizes some major reasons that VR is an excellent tool for education: first, it helps transform the abstract into the tangible, which is extremely useful in learning of abstract concepts like mathematics and geometry. Second, it provides hands-on practice opportunities by allowing a learner to learn from doing than just watching, which is very useful in areas such as medical training. Third, it makes it possible to situate learning experiences in places that are impossible or hard to reach. For example, Google Earth VR allows students to go anywhere on the earth and learn "in situ." Fourth, it breaks the boundaries of the physical world and allows a learner to explore learning with physically impossible and novel experiences. For example, students from a physics class can watch subatomic particle collision events in a supercollider VR simulator.

Studies on the effectiveness of using VR for learning have found that VR can promote learning by bringing positive learning experiences to learners and increase their engagement in learning. Merchant et al. [6] compared K-12 students' learning outcomes in three major categories of desktop VR applications: games, simulations, and virtual worlds. They found that all of the three types of desktop VRs are effective in improving learning to some degree, and that games in particular show higher learning benefits than simulations and virtual worlds. In another study on using immersive VR for mathematics and geometry learning, Kaufmann et al.[7] found that VR provided benefits to the learners by training their spatial ability and helping them solve simple mathematics and geometry problems. Using VR for science education has been a favorite of educators due to its powerful visualization and interactive features. ScienceSpace, an immersive VR application developed by Dede et al. [2] was a seminal project in this field. They built three immersive virtual worlds that allowed students to explore the topics of Newtonian mechanics, electrostatics, and molecular structure. Their study results showed that students enjoyed their learning in the virtual worlds and developed a deeper understanding of abstract science concepts.

B. Definitions of Interactivity

Defining interactivity is difficult due to the many different perspectives and contexts it is associated with [8, 9]. Discussions of interactivity are popular in the fields of computer-mediated communication [10], interactive instructional multimedia [11], and human-computer interaction [8, 12]. We will focus our discussion on the latter two fields, because they are closely related to our field of study. Sims [11] defines interactivity in interactive instructional multimedia as functions mapping input given to the user by the system and the system actions in response. Rhodes and Azbell [13] propose the concept of levels of interactivity within the same context in terms of reactive, coactive, and proactive. Specifically, at a low, reactive level, the learner has little control of media content and structure; at a coactive level, the learner has control of media sequence, pace, and style; and at the highest level, proactive, the learner is able to control both media content and structure.

From a human-computer interaction perspective, Steuer [12] defines interactivity in VR as the degree "to which users can participate in modifying the form and content of a mediated environment in real time." Roussou et al. [14] defines interactivity in a virtual learning environment (VLE) as "the process that actively involves the learners both physically and intellectually." She further defines three levels of interactivity in VLE: spatial navigation is the lowest form of interactivity; manipulation of the environment counts as a basic medium level of interactivity; and the ability to modify system of operation is the top level of interactivity.

In addition, Ritter [15] proposes a design strategy for controlling interactive media that connects learning concepts with specific interactivity and user interfaces. He defines the strategy as conceptual interactivity and states that conceptual interactivity is "the attainment of a concept through an interactive experience that contains a psychological and physical correspondence between media content and the user interface for directing that content." We are interested in this definition because in our design experiment, we tried to design user actions that directly lead to learning concepts and create custom user interfaces that facilitate the actions.

One of the core concepts of above definitions of interactivity is the amount and types of controls the system grants a user. Based on this concept and Ritter's design strategy of interactivity in interactive media, we propose that in the context of interactive VR storytelling for learning, the interactive features of an VR experience should support embedded learning concepts through learners' actions. In particular, interactivity provided by the VR experience should give a learner freedom in navigation, the ability to select virtual objects and apply intended effects, and the ability to access specific learning information in the environment.

C. Interactivity and Learning

How does interactivity affect learning in a VR experience? Several studies specifically addressed this topic in the literature. Although those studies vary in their constructs of VR interactivity, they present positive evidence that interactivity in VR experiences benefits learning.

In a non-immersive VR study [16], interactivity was mainly investigated as co-located collaborative actions between three users. Three levels of interactivity (fully automated, single active player, and three collaborating players) were compared in the study, and quantitative results show that the three-player group scored significantly higher than the other two groups on collaborative tasks in the game, suggesting the higher level of interactivity between users resulted in a better understanding of the learning task.

In another study, Bailenson et al. [17] examined the effect of interactivity, specifically in the form of system feedback, on learning physical actions (Tai Chi) in VR. In the study, they compared learners' objective and subjective learning in an interactive VR condition (users are taught by a virtual teacher and able to view themselves in playback with different views) with a non-interactive video condition (users watch a Tai Chi video with no feedback available). Their results demonstrate an advantage of the interactive VR condition over the non-interactive video condition in both subjective ratings and objective improvements in learning from two experiments.

Additionally, Roussou et al. [14] investigated the effect of interactivity on conceptual learning in a CAVE-based virtual environment. Their constructs of VR interactivity include user exploration, action, and reaction of virtual objects in the VE. Three conditions with different levels of interactivity (fully interactive VR, guided VR, and non-VR) were created in their experiments and played by primary school students. Evidence from the study showed that the guided VR condition, in which activity was guided by a virtual robot, supported students' conceptual change the best among all conditions.

D. *Interactivity and Cognitive Load Theory*

Cognitive load theory (CLT) holds that a learner's capacity of working memory is limited and if a learning task requires too much capacity, learning will be hindered [18]. CLT provides important design guidelines to the design of interactive instructional media, because increased interactivity and learner control may add to a learner's extraneous cognitive load and leave less working memory for actual learning tasks. Several studies have specifically addressed the question of how interactivity in multimedia affects a learner's cognitive process and learning outcomes. Moreno and Valdez [4] found that students in an interactive learning group with extra interactivity of arranging learning materials performed more poorly than students in other multimedia groups. They further found that the negative effect of the interactivity was due to students' lack of time control and system feedback, suggesting that for novice learners, interactivity may not promote learning unless it is carefully integrated into the design of different feedback strategies. In another study on the effect of cognitive load on map learning, Knight and Tlauka [5] compared learning performance with interactive and passive map learning. Their result suggests that in spatial learning, interactivity may be detrimental for cognitively demanding tasks. Kalet et al. [19] compared medical students' clinical skills performance in abdominal exams after treatment of a computer-assisted learning module with three different interactivity levels on content manipulation: watch, click, and drag. Their results show that there was no significant learning difference in knowledge test across all three groups, but students in the click group, in which they had a mid-range level of interactivity on learning content manipulation, scored significantly higher in physical exams than those in the other two groups. These results inform our hypothesis that a moderate level of interactivity may be the "sweet spot" for learning in immersive VR.

III. Experiment

A. *Goals and Hypotheses*

To investigate how different levels of interactivity in a storytelling-focused VR experience affect learning outcomes, we designed an experiment based on an immersive VR storytelling experience that was designed to supplement immunology learning at a college level. We proposed three hypotheses:

1. Level of interactivity will significantly affect a user's learning gains.

2. Users will achieve the highest learning gains in the medium interactivity condition.

3. Increased levels of interactivity will significantly increase a user's engagement in the learning experience.

We elaborate on the details of the experiment in the following sections.

B. *VR Learning Experience*

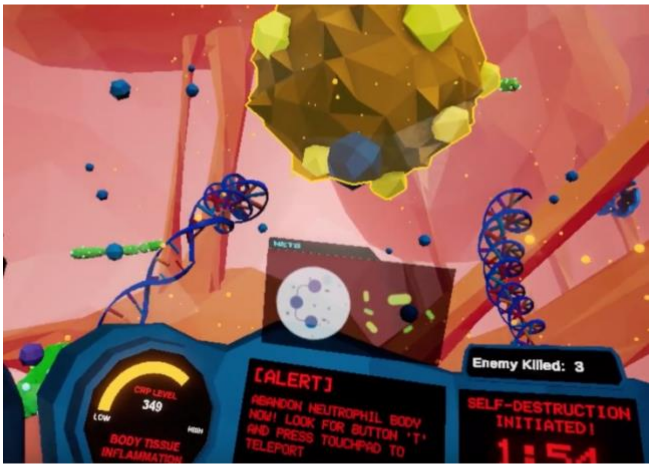

Fig. 1. Screenshot from the experience (User is using the NETS (neutrophil extracelluar traps) weapon to kill bacteria in body tissues)

We designed our immunology learning environment (Fig. 1) as a complete VR learning experience; it also serves a testbed to help with our investigation of the hypotheses above. We chose the topic of immunology for the experience because immunology concepts typically taught at the college level are inherently complex, abstract, and process-based, involving hard factual knowledge and jargon of the field [20] and presenting learning barriers to college students. Immunology knowledge can be difficult for students to visualize and relate to. Current methods of teaching immunology, usually in large-size didactic lecture classes plus lab exercises, do not promote learning in an active way for most students. The affordances of VR technology, such as powerful visualization and rich interactivity, make it a suitable instructional tool to convey the concept and bring a different learning experience to the students.

The virtual experience in our application focuses on a specific type of white blood cell called a neutrophil. Neutrophils travel through blood vessels until they are called to "transmigrate" to the site of infection inside body tissues. They use three major mechanisms (degranulation, phagocytosis, and neutrophil extracellular traps) to kill pathogens causing the infection, and each of the mechanisms has its own pros and cons.

To help students learn about neutrophils and their properties, our application puts the student into the role of a neutrophil "pilot," riding in the neutrophil as it patrols the blood vessel, "driving" the neutrophil to the transmigration portal by following chemical signals, navigating through the body tissue to get within range of enemy pathogens, and using their three "weapons" to do battle with the pathogens. Learners have the goal of defeating enough of the pathogens to restore their host's health to a target level while avoiding excessive inflammation. All along the way, a virtual guide helps the new “recruit” to learn how to use their neutrophil’s capabilities. This story builds on a classic conflict narrative, relatable metaphors, game-like objectives, and novel interactive mechanics to engage learners and facilitate learning. It also allows learners to visualize the invisible microscopic world inside the human body in a fun and novel way.

As shown in Figure 1, the learner sees a fully-surrounding computer-generated virtual environment. The neutrophil “cockpit” is visualized as a glass ball containing informational displays and weapons controls on a dashboard. Interactions include pointing at other cells to learn about them, using a controller to roll the neutrophil on the surface of the blood vessel, navigating in 3D inside the body tissues, choosing weapons, and firing weapons at pathogens. The virtual guide is presented via audio narration, and 3D models and animations help learners understand the processes used by neutrophils to fight off infections. More detail on the design of our application is provided in an earlier paper [1].

C. Experiment Design

1) Levels of interactivity

We wanted to study three conditions at low, medium, and high levels of interactivity. We were inspired by the levels of interactivity in existing digital media and interactive entertainment. For example, at the very low end of interactivity are the traditional TV and movie watching experiences, in which the audience has no options to interact with content. At the other end of the interactivity spectrum, many “open world” or “sandbox” video games allow players to interact in almost any way imaginable with virtual objects and environments. A medium level of interactivity could be anything between these extremes. We defined the amount of interactivity available to users in our medium interactivity condition by carefully combining system automation and interactive key actions (user actions directly related to the learning concepts). In short, our definitions of the levels of interactivity in the study are as follows:

- Low interactivity: System automates as many actions as possible, no user control beyond choosing where to look in the VE.
- Medium interactivity: A balance between system automation and user-controlled actions, where the interactive components are directly related to learning concepts.
- High interactivity: The user chooses and initiates as many actions as possible, while still maintaining narrative flow.

The experiment used a between-subjects design with level of interactivity as the independent variable. Each participant was randomly assigned to one of the three conditions (low, medium, or high).

We modified the original VR learning experience to create three versions of the experience with low, medium, and high levels of interactivity. We divided user actions in the experience into three categories: information access, travel, and interaction with virtual objects. Table 1 shows examples of user actions for each action category and how system automation and user-controlled actions were assigned in each action category at low, medium, and high levels of interactivity.

To access information about immune cells in the virtual blood vessel (left column of Table 1), learners in the low interactivity condition see information pop up automatically when the neutrophil gets close to the cells. In the medium interactivity condition, a cell’s information appears when it is nearby and the learner is looking at it. Finally, in the high interactivity condition, the learner has to actively select a specific cell using a laser beam in order to see its information.

There were two different types of travel (middle column of Table 1). In Scene I, the neutrophil must roll along the blood vessel wall to reach the portal, while in Scene II, the neutrophil has to move around in the tissue to get within range of the pathogens. Both of these were automated in the low interactivity condition, and both were fully manual in the high interactivity condition. At the medium interactivity level, Scene I travel was under user control (since searching for the portal is an important learning concept), while Scene II travel was automated (since moving within range of pathogens is not tied to critical learning concepts). However, in order to provide some user control in Scene II, the medium interactivity level did allow learners to turn the neutrophil in different directions to target specific pathogens.

Using the weapons to kill pathogens (right column of Table 1) involved selecting a weapon and firing it at a nearby pathogen. These processes were automated in the low interactivity condition, and fully under user control in both the medium and high interactivity conditions.

To achieve experimental control, the three conditions were as similar as possible for other elements of the experience beyond interactivity. All conditions told the same story and presented the same information related to the immunology concepts. Of course, details of the experience (e.g., how many pathogens learners interacted with) might vary based on user choices in the more interactive conditions.

TABLE 1. ASSIGNMENT OF ACTIONS TO THE SYSTEM AND USER IN LOW, MEDIUM, AND HIGH INTERACTIVITY CONDITIONS

	Information Access	**Travel**	**Interaction with Virtual Objects**
	Example: Immune cell information and weapon use how-tos	***Example:*** Movement in blood vessels and body tissues	***Example:*** Killing of bacteria using different weapons
Low interactivity	System-automated	System-automated	System-automated
Medium interactivity	System + User	System+User	User-controlled
High interactivity	User-controlled	User-controlled	User-controlled

2) Measures

Objective measure. We used a pre-test and a post-test before and after the VR experience to measure study participants’ learning gains. Each of the two tests had the same 15 items in the form of multiple-choice and matching questions.

The order of the choices were shuffled on the post-test in order to reduce possibility of memorization of the answers. The tests were scored on a 15-point scale, with each item being worth one point.

Subjective measures. We used three questionnaires and a short interview to collect participants' feedback on learning and user experiences. Specifically, we used the User Experience Questionnaire (UEQ) [21] to measure both classical usability aspects (efficiency, perspicuity, dependability) and user experience aspects (originality, stimulation) of our interactive VR application. Because we embedded some gaming elements in our VR experience design, we used a core module of the Game Experience Questionnaire (GEQ) [22] to assess the participants' game experience on seven components: Immersion, Flow, Competence, Positive and Negative Affect, Tension, and Challenge. Additionally, we developed our own learning experience questionnaire (LEQ) to measure the participants' subjective learning experience directly related to design elements embedded in the experience. The questionnaire asked about participants' learning experience, specifically their perceptions of: comprehension of embedded learning concepts; effectiveness of VR for concept comprehension; engagement with the experience; aesthetics of the environment; and effectiveness of the audio guide, interaction, animation, and system feedback on concept learning. For each question, participants answered on a scale ranging from 1 to 5, with higher numbers indicating more positive responses. We also used a short interview to collect more information on general user playing and learning experiences. Specifically, we asked:

- What do you think of learning immunology through a virtual reality experience?
- How would you compare learning immunology through virtual reality to normal classroom learning?
- What part of this virtual experience did you find the most useful and interesting for learning?
- What was the most challenging/difficult part of this experience during your learning?
- How could this experience be improved to better help with learning immunology concepts?

3) Apparatus

The VR experience was designed in the Unity5 game engine and was displayed in an HTC Vive Pro VR system, which includes a high-resolution headset, two six-degree-of-freedom controllers, and two Lighthouse trackers. We chose the Vive system due to its higher headset resolution for sharper graphics, more available development tools, and ergonomic design that allows users with glasses to use it comfortably. Visual assets for the experience were custom designed to represent the key features of the immunology concepts, but in a non-photorealistic low-polygon style that is reminiscent of popular video games. Since the user was seated throughout the experience, we used a high-end desktop computer to run the experience in the game engine.

4) Participants and Procedure

A total of 62 participants (42 females, 20 males, mean age 19 years old) were recruited for the study; they were all from a freshman level biology orientation class in the Department of Biological Sciences at a large research university. All participants had normal vision (with or without correction). Most participants had no prior experience with VR (26 people), or only a small amount of prior VR experience (35 people). Only one participant had a moderate amount of VR experience previously.

Participants provided informed consent and completed the pre-test online, and were asked to do so at least 24 hours before their scheduled study session to lessen the possibility that they would remember the questions while using the VR application. Three participants took the pretest at the site of study since they forgot to do so beforehand. On the day of the study, each participant was assigned to one of the three experimental conditions and given the corresponding experience prototype to play. We assigned 20 participants to the low interactivity group, 21 to the medium interactivity group, and 21 to the high interactivity group. Immediately after the experience, the participant was given the post-test and an online survey to fill out. The survey contained demographic questions, the LEQ, the UEQ, and the GEQ. After the participant completed the survey, the short interview concluded the study.

D. Results

1) Objective Learning Results

The SPSS Statistics software package was used to conduct all analyses of quantitative data collected in the study. Two outliers from test scores (one in the low interactivity group and one in the medium interactivity group) were removed from further data analysis because they were more than two standard deviations away from the test score means. We first did a paired sample T-test (Fig. 2) on test scores in all conditions together and found a significant difference ($t(58) = -7.3$, $p < 0.0005$) between pre-test scores ($M = 8.24$, $SD = 2.18$) and post-test scores ($M = 10.51$, $SD = 2.12$), which indicates that the VR experience did result in participants learning immunology concepts to some extent. We then calculated learning gains for each participant by subtracting their pre-test score from their post-test score. Figure 3 presents boxplots of the learning gains in each condition. We ran a one-way between-subjects ANOVA to compare learning gains between the three different conditions. Our analysis did not find any significant effect of interactivity on learning gains. As seen in Fig. 3, the median learning gain was similar in all three conditions, and the variance was large.

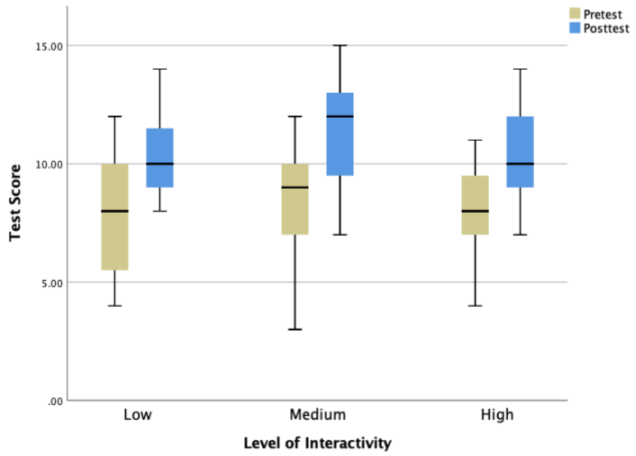

Fig. 2. Pre- and post-test scores for all conditions together (excluding two outliers). The boxplots indicate the median values (black lines), 2nd and 3rd quartiles (tan and blue boxes), and lowest and highest test scores (whiskers).

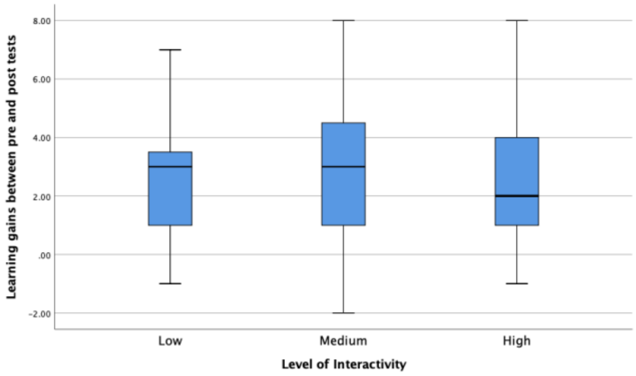

Fig. 3. Learning gains in the three conditions (excluding two outliers).

2) Subjective learning results

We used Kruskal-Wallis tests to analyze the non-parametric data collected from the three questionnaires. The GEQ contains seven sub-components (competence, sensory and imaginative immersion, flow, tension/annoyance, challenge, negative affect, positive affect). Our analysis revealed a significant effect of interactivity on challenge (χ^2 (2) = 6.419, p = 0.04). Post-hoc pairwise comparisons showed a significant difference between the low and medium interactivity conditions (t = -14.031, p = 0.036), with the medium level of interactivity being perceived as more challenging than the low interactivity condition (Fig. 4). There were no significant effects of interactivity found on other GEQ sub-components.

The UEQ measures six attributes of a product experience (attractiveness, perspicuity, efficiency, dependability, stimulation, novelty). Our analysis revealed no significant effects of interactivity level on any of them. Using the ranges supplied by the authors of the UEQ [21], we found that all three conditions were rated in the 'good' to 'excellent' categories in all six aspects of user experience. The fact that UEQ scores were uniformly high perhaps created a ceiling effect that made it difficult to detect any differences between the condition based on this measure.

On the LEQ, significant effects of interactivity were found in four categories:

- Effectiveness of VR learning experience (χ^2 (2) = 6.107, p = 0.047). A post-hoc test revealed that the medium interactivity level was perceived as significantly more effective than the low interactivity level (t = -12.445, p = 0.05).
- Engagement of VR learning experience (Fig. 5) (χ^2 (2) = 11.373, p = 0.003). A post-hoc test revealed that the high and medium interactivity levels were perceived as significantly more engaging than the low interactivity level (high and low, t = -11.736, p = 0.01; medium and low, t = -11.736, p = 0.01).
- Effectiveness of interacting with virtual environment for learning (χ^2 (2) = 14.956, p = 0.001). A post-hoc test revealed that the interaction in the high interactivity level was perceived as significantly more effective for learning than the low interactivity level (t = -17.075, p < 0.0005).
- Effectiveness of animation for learning (χ^2 (2) = 7.203, p = 0.027). A post-hoc test revealed that the animation in the medium interactivity level was perceived as significantly more effective for learning than the low interactivity level (t = -11.726, p = 0.035).

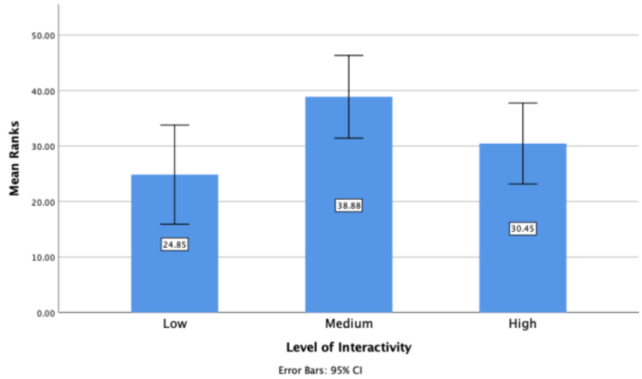

Fig. 4. Differences between Challenge scores (mean rank) in GEQ

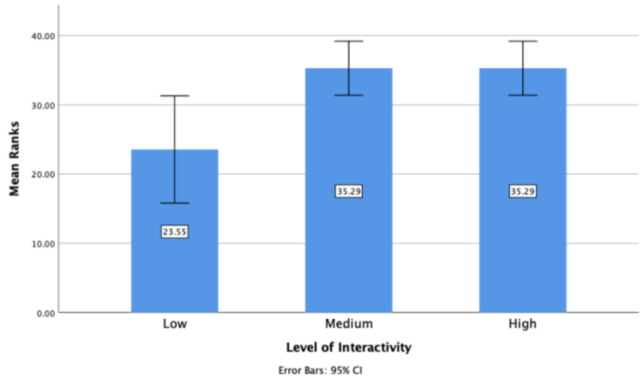

Fig. 5. Differences between VR learning engagement scores (mean rank) in LEQ

IV. Discussion

Recall that our hypotheses were:

1. Level of interactivity will significantly affect a user's learning gains.
2. Users will achieve the most learning gains in the medium interactivity condition.
3. Increased levels of interactivity will significantly increase a user's engagement in the learning experience.

Regarding hypotheses 1 and 2, our analysis of the objective learning gains data did not support these hypotheses. We found no measurable effect of interactivity on learning gains. This of course does not mean that interactivity is irrelevant to learning in a VR storytelling experience, just that we were not able to detect any such effect in our experiment. Many, if not most, studies that compare different educational materials or media are unable to detect direct effects on learning as measured by pre- and post-tests [23, 24], perhaps because learning gains by individual students are influenced by individual differences and many other factors outside the materials being studied. For example, prior knowledge, working memory size, and information-processing patterns all vary from person to person. Cox [25] argues that learning is affected by a complex interaction of representations' properties, task demands and within-subjects factors.

Regarding hypothesis 3, our analysis of the subjective data revealed that the hypothesis is supported by both our LEQ data and participants' subjective perceptions from short interviews.

The LEQ analysis showed that learners felt significantly more engaged in learning in both the medium and high interactivity conditions compared to the low interactivity condition. Analysis of the interviews also supports this claim. Specifically, 11 participants (unprompted) in the medium and high conditions indicated they felt highly engaged, as opposed to only three in the low interactivity condition. Further, nine participants in the medium and high conditions mentioned that they felt focused on learning content, as opposed to two in the low condition. This finding is in accordance with some results from previous VR learning studies that increased interactivity will lead to increased engagement in learning [16]. Since the positive impact of engagement on learning is well known [26], this is a positive sign that providing more than a minimal level of interactivity in a narrative-based VR educational experience could be beneficial for learning.

Beyond our initial hypotheses, we uncovered other noteworthy findings. The GEQ analysis indicated that participants felt significantly more challenged in the medium interactivity condition than in the low condition. The challenge component in the GEQ is calculated from responses to five statements: *I thought it was hard, I felt pressured, I felt challenged, I felt time pressure, I had to put a lot of effort into it*. Thus, challenge in the GEQ seems to be defined as a neutral element (neither fully positive nor fully negative). The effect of interactivity on the GEQ's negative affect component was nearly significant (χ^2 (2) = 5.283, p = 0.07), with the low interactivity group having a much higher average negative affect value than the other two conditions (M_{low} = 38.63, M_{medium} = 28.31, M_{high} = 27.90). Since increased level of interactivity led to significantly more challenge but less negative affect, we tentatively interpret the increased level of challenge due to interactivity as a positive sign that may contribute to learning. Both the learning and game research literature indicates that an appropriate amount of challenge can lead to better performance on tasks that requiring integration and synthesis of concepts [27], and that it can add enjoyment and fun to gameplay [28].

The LEQ analysis showed that participants perceived the medium interactivity condition to be significantly more effective in helping them understand immunology concepts compared to the low interactivity condition. Participants also perceived that a high level of interactivity with virtual objects in the VE was significantly more effective for their learning in VR than a low level of interactivity. Finally, participants believed that interactivity-triggered animations in the medium condition were significantly more effective than automatically triggered animations in helping them understand embedded immunology concepts. All of these findings show that participants viewed interactivity as useful and positive in a learning context. Even though participants were not aware that there were three different levels of interactivity, they still provided significantly higher ratings for the medium and high conditions than for the low condition. While the low interactivity condition is similar to watching a 360-degree video, we observed participants becoming bored and distracted in this condition, while the higher levels of interactivity gave participants more choices and opportunities to engage with the learning content.

Based on these findings, we propose several tentative design guidelines for interactivity design in a storytelling-driven VR experience for concept learning:

- Use interactivity to enhance engagement with the content, but do not feel compelled to use interactivity for every possible action. A combination of user control and automation will often be appropriate.

- Use interactivity for key actions that are directly related to learning concepts. Designers should provide interactive components that allow users to put knowledge to use, where both successes and failures help reinforce the learning concepts.

- Try to design interactivity that leads directly to concept visualization or animation for a more straightforward learning experience that reduces unnecessary cognitive load.

V. Limitations

Several limitations of our study are worth noting. First, our user test sample is relatively small (only 20+ participants for each condition) and more participants may be needed to see statistically significant differences in objective learning outcomes. Second, the test items used for our pre- and post-tests were not tested for their validity as a measure of concept knowledge. Therefore, we do not know if they have correctly measured participants' learning outcomes as intended. Validation of test items used for objective measure of learning is needed for future experiments. Third, although we carefully designed and planned most key actions that directly connect to learning concepts, there were some actions that did not lead to the learning concepts as we intended and became extraneous interactivity that may hinder learning experience. For example, in the high interactivity condition, we gave the users full control of their virtual movement (rotation and translation) during the process of searching for the portal to transmigrate. However, some users found the virtual movement hard to control, and thus felt disoriented and spent quite some time to reach the portal. Rather than helping them learn about neutrophil movement and transmigration, this frustrated and distracted users from the learning experience.

VI. Conclusions and Future Work

In this study, we present our work on an investigation of the level of interactivity and its effect on learning of immunology concepts. Although we did not find evidence from objective learning data to support our hypotheses that the level of interactivity will affect learning outcomes, we found strong evidence from subjective data that support our hypothesis that increased level of interactivity will lead to increased engagement in VR learning. We also had other findings that could inform the interactivity design in a storytelling-based VR experience for learning. Specifically, our subjective learning data suggest that a higher level of user interaction with virtual objects in the VE will lead to more effective concept learning, and that interactivity-triggered concept visualization and animation are more effective for learning than automated visualization and animation.

Our experiment in the study is just the first step of our research agenda to systematically investigate the impact of interactivity on learning in storytelling-based VR applications. We hope to be able to further investigate various aspects of interactivity design and its use for learning difficult concepts—not just how much interactivity to be used in a VR learning experience, but how the interactivity can be connected to concept learning more intuitively based on its types and properties to enhance the learning experience. Besides studies of interactivity, we are also interested in how different aspects and elements of storytelling impact learning in VR and hope to conduct relevant studies in future.

REFERENCES

[1] L. Zhang, and D. A. Bowman, “Enabling immunology learning in virtual reality through storytelling and interactivity”, in Proceedings of Human-Computer Interaction International (HCII) Conference 2019, in press.

[2] C. Dede, M. C. Salzman, R. B. Loftin, and K. Ash, “Using virtual reality technology to convey abstract scientific concepts”, in Learning the Sciences of the 21st Century: Research, Design, and Implementing Advanced Technology Learning Environments. London: Lawrence Erlbaum, 1997.

[3] M. Slater, MV. Sanchez-Vives, “Enhancing our lives with immersive virtual reality,” Frontiers in Robotics and AI. Vol.3. 2016.

[4] R. Moreno, & A. Valdez, “Cognitive load and learning effects of having students organize pictures and words in multimedia environments: The role of student interactivity and feedback”, Educational Technology Research and Development, 2005, 53, pp.35–45.

[5] M. J. Knight and M. Tlauka, “Interactivity in map learning: the effect of cognitive load”, Spatial Cognition & Computation, 2016, 17(3), pp.1-14

[6] Z. Merchant, E.T. Goetz, L. Cifuentes, W. Keeney-Kennicutt, T. J. Davis, “Effectiveness of virtual reality-based instruction on students’ learning outcomes in K-12 and higher education: a meta-analysis”, Computer Education. 2014, 70. pp.29–40.

[7] H. Kaufmann, D. Schmalstieg, and M. Wagner, “Construct3D: a virtual reality application for mathematics and geometry education”, Education and Information Technologies, December 2000, Volume 5, Issue 4, pp. 263–276.

[8] M. Roussou, “Can interactivity in virtual environments enable conceptual learning?,” (Proceedings) 7th Virtual Reality International Conference (VRIC), 2005, First International VR-Learning Seminar, pp. 57-64, Laval, France.

[9] S. Kiousis, “Interactivity: a concept explication,” New Media Society, 2002, 4, 355

[10] E. J. Downes, S. J. McMillan, “Defining interactivity: a qualitative identification of key dimensions”, in New Media & Society, 2000, volume 2, issue 2, pp. 157-179

[11] R. Sims, “Interactivity: A forgotten art?” Computers in Human Behavior, 1997, volume 13, issue 2, pp. 157-180.

[12] J. Steuer, “Defining virtual reality: dimensions determining telepresence,” Journal of Communication, 1992, volume 42, issue 4, pp. 73-93

[13] D. M. Rhodes and J. W. Azbell, “Designing interactive video instruction professionally”, Training and Development Journal, 1985, 39(12), pp.31-33.

[14] M. Roussou, M. Oliver, and M. Slater, “The virtual playground: an educational virtual reality environment for evaluating interactivity and conceptual learning,” Virtual Reality, 2006, 10, pp.227–240

[15] D. Ritter, “Defining conceptual interactivity,” 2014 International Conference on Virtual Systems & Multimedia (VSMM), 2014.

[16] P. Apostolellis, and D. A. Bowman, “C-OLiVE: Group co-located interaction in VEs for contextual learning,” in Proceedings of the IEEE International Conference on Virtual Reality (VR), pp. 129-130. Minneapolis, MN, USA, 2014.

[17] J. Bailenson, K. Patel, A. Nielsen, R. Bajscy, S-H. Jung, G. Kurillo, “The effect of interactivity on learning physical actions in virtual reality”, Media Psychology, 2008, 11:3. 354-376.

[18] T. D. Jong, “Cognitive load theory, educational research, and instructional design: some food for thought,” Instructional Science, 2010, volume 38, issue 2, pp. 105-134

[19] A. L. Kalet, HS. Song, U. Sarpel, R. Schwartz, J. Brenner, T. K. Ark, J. Plass, “Just enough, but not too much interactivity leads to better clinical skills performance after a computer-assisted learning module,” Med Teach. 2012, 34, pp. 833-839.

[20] S. L. Raimondi, “ImmuneQuest: Assessment of a video game as a supplement to an undergraduate immunology course,” Journal of Microbiology & Biology Education, May 2016, pp. 237-245.

[21] M. Schrepp, “User experience questionnaire handbook: All you need to know to apply the ueq successfully in your projects,” 2015.

[22] W. A. IJsselsteijn, Y. A. W. de Kort, and K. Poels, “The game experience questionnaire,” Eindhoven: Technische Universiteit Eindhoven, 2013.

[23] R. E. Clark, “Reconsidering research on learning from media”, in Review of educational research, 1983, 53(4), pp. 445-459.

[24] R. E. Clark, and T. G. Craig, “Research and theory on multimedia learning effects”, in Interactive multimedia learning environments, 1992, pp. 19-30, Springer, Berlin, Heidelberg.

[25] R. Cox, “Representation construction, externalised cognition and individual difference,” Learning and Instruction, 1999, 9 (4), pp. 342-363.

[26] R. M. Carini, G. D. Kuh, S. P. Klein, “Student engagement and student learning: testing the linkages”, in Research in Higher Education, 2006, volume 47, issue 1, pp. 1-32,

[27] T. K. O’Mahony, N. J. Vye, J. D. Bransford, R. Stevens, E. Sanders, R. Stephens, and M. K. Soleiman, “A comparison of lecture-based and challenge-based learning in a workplace setting: Course designs, patterns of interactivity, and learning outcomes”, Journal of the Learning Sciences, 2012, 21, pp.182–206.

[28] R. J. Pagulayan, K. Keeker, T. Fuller, D. Wixon, R. L. Romero, and D. V. Gunn, “User-centered design in dames”, The human-computer interaction handbook, 2002, pp. 883-906.