Exploring Effect of Level of Storytelling Richness on Science Learning in Interactive and Immersive Virtual Reality

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

Doug A. Bowman
Center for Human-Computer Interaction, Dept. of Computer Science, Virginia Tech
dbowman@vt.edu

ABSTRACT
Immersive and interactive storytelling in virtual reality (VR) is an emerging creative practice that has been thriving in recent years. Educational applications using immersive VR storytelling to explain complex science concepts have very promising pedagogical benefits because on the one hand, storytelling breaks down the complexity of science concepts by bridging them to people’s everyday experiences and familiar cognitive models, and on the other hand, the learning process is further reinforced through rich interactivity afforded by the VR experiences. However, it is unclear how different amounts of storytelling in an interactive VR storytelling experience may affect learning outcomes due to a paucity of literature on educational VR storytelling research. This preliminary study aims to add to the literature through an exploration of variations in the designs of essential storytelling elements in educational VR storytelling experiences and their impact on the learning of complex immunology concepts.

CCS CONCEPTS
• Human-centered computing → Human computer interaction (HCI); Interaction paradigms: Virtual reality; Human computer interaction (HCI); Empirical studies in HCI. • Applied computing → Education; Interactive learning environments.

KEYWORDS
Virtual Reality, Immersive Storytelling, Science Education, Human-Computer Interaction

ACM Reference Format:

1 INTRODUCTION
Immersive and interactive storytelling in virtual reality (VR) is an emerging creative practice that has been thriving in recent years. Educational applications using immersive VR storytelling to explain complex science concepts have very promising pedagogical benefits because on the one hand, storytelling breaks down the complexity of science concepts by bridging them to people’s everyday experiences and familiar cognitive models, and on the other hand, the learning process is further reinforced through rich interactivity afforded by the VR experiences. However, it is unclear how different amounts of storytelling in an interactive VR storytelling experience may affect learning outcomes due to a paucity of literature on educational VR storytelling research. Specifically, to create a story related to or based on specific learning concepts, how much detail should a story creator put into the essential elements of a story (e.g., the number of characters), so that the story will effectively convey the intended information to learners? Can a dramatic story help learners retain the information longer than a simple, more basic story? Those are the questions left unanswered by existing studies of storytelling and learning, and they inspired our research design of this study. We choose immersive and interactive virtual reality (VR) as a medium to carry learning-concept-driven stories for the study because its unique affordances, such as presence, high interactivity, and first person POV, help contribute to a greater impact on immersion and user experience than any other media. Since a compelling story has the potential to help learners connect more with its content and increase knowledge retention [1, 2], we are very interested in how to build a rich storytelling experience in immersive VR based upon specific learning concepts from immunology and how it impacts on a learner’s learning experience. In this paper, we present our preliminary research on how different storytelling element designs, in the form of level of storytelling richness, affect learning of immunology concepts in immersive and interactive VR experiences.

For the purposes of this study, we define certain important terms below:
• Storytelling elements: the key components of a story, which include plot, story structure, characters, dialogue, conflict, resolution, and point of view, etc.
• Level of storytelling richness: the amount of detail, nuance, and drama used for the storytelling elements in a particular story.

Results from our preliminary research showed although there was no evidence of significant differences found between different experiment conditions from objective and subjective quantitative measures, there was evidence from user interviews that higher levels of storytelling richness was associated with positive emotions and empathy from more learners and appeared to be effective in helping some learners retain learning information and build their confidence.

Our primary contributions are:
• We provide design guidance on storytelling designs in educational VR experiences with interactivity.
• We provide empirical evidence of how different levels of storytelling richness in essential storytelling elements affect user experience and learning in educational VR.
• We share our unique level of richness-based design process of creating engaging and custom stories from specific science concepts for educational VR experiences.

2 RELATED WORK

2.1 VR Applications in Science Education

Using VR in science education to promote learning has been in abundant practice since the 1990s due to its powerful visualization and interactive features. ScienceSpace, an immersive VR application developed by Dede et al. [3] allowed students to explore the topics of Newtonian mechanics, electrostatics, and molecular structure in three different virtual worlds. Their study results showed that students enjoyed learning in the virtual worlds and developed a deeper understanding of abstract science concepts. The Construct 3D project by Kaufmann et al. [4] explored VR application in mathematics and geometry education with an aim to train students’ spatial capabilities. Their study showed the usefulness of the system in helping the students with their understanding of complex spatial problems and relationships in 3D spaces. Manseur [5] also explored a desktop VR tool, the Virtual Reality Modeling Language (VRML), in chemistry and robotics education and demonstrated its great potential in generating interactive and dynamic computer graphics to help students understand scientific and technological principles. These practices and studies of using VR in science education have made VR technology integration in science learning design a very promising practice.

2.2 Using Storytelling to Promote Science Learning

Storytelling has been adopted widely by educators as a pedagogical tool to help convey complex information to learners. According to Abrahamson [6], “storytelling is an important technique in the process of learning and understanding” from a cognitive processing perspective. It uses concrete examples that have relationships to life experiences rather than vague abstractions and generalizations in communicating information. Emotions carried by the stories, such as inspiration, encouragement, satisfaction, and fascination, weaving together with information, can also have an impact on students’ long-term memories [6].

The advantages of using narrative text to convey information over expository text make it promising to use stories to explain hard scientific concepts for better understanding. Educators have used stories in science teaching to make complex science subjects accessible to novice learners with positive results [7–9]. Helstrand and Ott [7] used a science fiction novel, The Time and Space of Uncle Albert, to teach the theory of relativity in four classes. Their results indicate that using novels to teach scientific theories is an efficient way to help students acquire basic science concepts. In another study, Banister and Ryan [8] developed a science-teaching pedagogy using storytelling to help children develop science concepts about the water cycle. Their outcomes showed that, in the long run, the children remembered more abstract science concepts when taught with a storytelling-based pedagogy. RiverCity [9] is another research project that uses digital storytelling to train middle-school students in scientific inquiry skills in an interactive multi-user virtual environment. The project is backed by a strong storyline in which a student travels back in time, bringing 21st century technology and skills to investigate and solve 19th century health problems in a small town. Results suggest that interactive storytelling in a multi-user virtual environment effectively encourages authentic scientific inquiry in middle school science education.

Evidence from these studies suggests the possibility of using custom designed stories in immunology education to break down the complexity of abstract concepts and communicate them in an easy way to students, which is a major goal of our current study.

2.3 Story Structure and Type Related to Effective Learning

Strategies for creating effective stories to support learning have been explored and discussed extensively in literature [10–13]. Many try to connect storytelling with models and frameworks in learning sciences to optimize its pedagogical potentials [11, 13]. However, we are specifically interested in how to vary designs in essential storytelling elements and tweak their level of richness to promote learning. Literature on this specific topic is very limited and we drew some inspiration from following two studies of interest. O’Neill [12] developed a story typology that utilizes different story types for instructional purposes in an organizational setting and suggested different story types may have different instructional effectiveness. In the typology (Figure 1), color is defined as “structural elements, including lyricism, picturesque description, and vivid detail; and plot content such as heroic, comedic, and/or romantic elements” [12]. The typology shows differences between four story types. Specifically, epic stories with high color and high need fulfillment make an effective story structure for instructional purposes, meeting both storyteller’s and listeners’ needs.

Ferguson, et al [14] investigated the role of interaction mode and story structure in VR serious games and compared how an explicit strong story structure affects learning versus an implicit story structure. Their results showed that the explicit story structure and a guided experience led to higher retention of story content, such as factual information and knowledge, while the implicit story structure only helped with retention of more spatial information in the virtual world.

2.4 Immersive Storytelling in VR

Compared with traditional ways of storytelling, immersive storytelling in VR has unique advantages and is booming in recent years. Bosworth and Sarah [15] did extensive interviews with practitioners of immersive storytelling and found that storytelling mediated by VR technologies allows people to gain exclusive access to places where they couldn’t go in normal life, helps close distance between viewers and artifacts and make the story more relevant, and builds empathy in the audience and makes the experience more real by putting them in someone else’s shoes or environments. However, though with great instructional potentials, a major drawback in many of such immersive storytelling experiences, mostly in the
form of cinematic VR, is that they only provide limited interactivity
to their users and learning through the experiences tends to be
passive. We hypothesize that both interactivity and storytelling are
two essential components of an educational VR experience that
can contribute to positive learning outcomes. Therefore, we aim
to design immersive storytelling experiences that are also highly
interactive for our current research.

3 EXPERIMENT

3.1 Goals and Hypotheses

The overarching goal of this study is to explore how different story-
telling designs, especially designs that vary the richness of essential
storytelling elements such as plot, characters, dialogue, setting, con-
flict & resolution, and voice, affect a user’s learning experience in
an immersive and interactive educational VR experience for im-
munology learning. Based on our research interest, we came up
with four hypotheses.

- H1. Level of storytelling richness will significantly affect a
  learner’s learning gains.
- H2. Stories with higher levels of richness will engage learners
  more in an educational VR experience.
- H3. The learners in the advanced level of storytelling rich-
  ness group will have the highest rate of retention of learning
  information one week after the VR experience.
- H4. Compared to participants in the lower level of story-
  telling richness condition, more participants in higher level
  of storytelling conditions will perceive storytelling elements
  as useful factors for effective learning.

3.2 Immunology VR Learning Environment

We designed Immunology VR, a fully immersive immunology learn-
ing virtual environment (Figure 2), as a testbed to help with our in-
vestigation of research hypotheses for interactivity and storytelling.
We chose a challenging immunology topic, neutrophil transmi-
gation and killing mechanisms, identified by professors and students
in the Biological Sciences Department at a large research univer-
sity. The concepts covered by the topic are inherently complex,
abstract, and process-based, involving hard factual knowledge and
terminology of the field, which are therefore difficult for students
to visualize and understand. The VR application puts a learner 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.

3.3 Define the Scope of Immersive VR

Storytelling in the Current Study

We focused the scope of our storytelling research on the computer-
generated interactive and immersive VR experiences driven by a
dramatic story structure, an experimental combination that fea-
tures characteristics of cinema, drama, immersive storytelling, and
interactive virtual reality. The story from the experience is devel-
oped from specific learning concepts chosen from immunology and
its development will follow a three-act storytelling model with a
strong focus on character development. In short, our immersive
storytelling experience will be first-person POV, built upon a fully
computer-generated cinematic VR with abundant interactive agents
that directly connect to embedded learning concepts.

3.4 Storytelling Design

3.4.1 Elements of Storytelling. To create stories and tweak their
contents for specific needs, one has to be familiar with basic sto-
rytelling elements. Therefore, our first step is to decide what are
some of the common essential elements in storytelling and which of
them can be easily varied to meet our research needs. We compared
several different resources on basic storytelling elements (Table 1)
and decided to choose the most commonly mentioned storytelling
elements (plot, characters, setting, conflict, resolution), as well as
some that are suitable for learning design purposes in our project
(discussion and point of view).

3.4.2 Story Structure. Although storytelling is a frequently used
design element in many educational applications, stories created
from it differ a lot in their structures. To create a short (20-25
minutes), dramatic story with embedded immunology concepts, we
looked into traditional storytelling models for films and plays and
adopted them into a story structure that met our instructional needs.
We used screenplay-style writing as a main format to compose the
story script due to its flexibility with rich character development and abundant dialogue. Aristotle’s classic three-act story arc [16] and Syd Field’s extended Aristotle model [17] for screenplay are the main models we referenced when developing our own story structure. Figure 3 illustrates how our story structure maps onto Field’s screenplay structure.

Referencing the three act structure for film, we developed the following synopsis for the VR experience:

The story is set inside the human body in the blood vessel and the body tissues surround it. PHIL, the main character embodied by the user, is a young neutrophil just dispatched to a site of infection caused by S. Aureus bacteria (supporting characters). PHIL meets its friend MAC (supporting character), a macrophage at the site of the infection and learns how to use its three unique killing mechanisms (degranulation, phagocytosis, NETS) to fight the bacteria. After killing several S. Aureus, the biggest conflict of the story comes when PHIL detects a dangerous mutant of S. AUREUS (the main antagonist) and decides to trace it down and stop it from growing and spreading in the host’s body. However, after a brief confrontation, PHIL discovers that the MUTANT is very resistant to its common killing mechanisms, and what makes the situation worse is that the MUTANT is new to the immune system and its information hasn’t been registered. Determined to eliminate the MUTANT from the host’s body by any means, PHIL comes to a final resolution by activating its deadly but suicidal weapon, NETS, to trap the MUTANT with its DNA substances and break it apart. In doing so, it allows MAC, an antigen presenting cell, to successfully process the MUTANT’s antigen information and pass it on to many other immune cells, thus activating a full-body immune response against the MUTANT and its copies in the human body.

3.4.3 Characters. Characters are important storytelling elements in the development of a story arc. In detailing a character for screenplays, Field notes that a good character needs to have four essential qualities: 1. Has a strong and defined dramatic need; 2. Has an individual point of view; 3. Personifies an attitude; 4. Goes through
some kind of change, or transformation [17]. Those essential qualities of a character provide a great way to add layers of detail when creating a character for an advanced story.

Based on the learning concepts, we developed several characters (Figure 4) for the advanced level of storytelling with application of Field’s four essential qualities of a good character to our main character (PHIL) design.

Protagonist:
- PHIL: a young neutrophil, embodied by the user.

In the story, PHIL is born to be an immune cell foot soldier whose paramount duty is to obey the orders from the command center and defend its host’s body. It has a strong will to kill pathogens by any means. As a young neutrophil, PHIL is a highly-motivated learner who is eager to master all of its killing mechanisms (1. *Has a strong and defined dramatic need*). PHIL’s life goal is to kill as many bacteria as possible within its short life (2. *Has an individual point of view*). Warm-hearted and friendly by nature, PHIL is merciless to its enemy, bacteria that threaten its host’s body (3. *Personifies an attitude*). Although PHIL enjoyed its extended life after transmigration from the blood vessel to the body tissues, its sense of duty as an immune soldier causes it to make a very tough decision (4. *Goes through some kind of change, or transformation*) by activating its suicidal weapon, NETS, when fighting a bacterium MUTANT that cannot be killed by common killing mechanisms.

Antagonist:
- THE MUTANT: a mutated bacterium that is hard to kill

Other characters:
- MONO6/MAC: a monocyte who later differentiates into a macrophage, and a companion to PHIL
- MAJOR 16: a memory T Cell that recruits other white blood cells to the infection site
- ARA-C: PHIL’s virtual assistant
- S.AUREUS: a bacterium that causes infection
3.4.4 Dialogue. Dialogue among the characters plays another indispensable role in our VR learning experience design. Story dialogues serve several major purposes, such as advancing the plot of a narrative, providing information, exposing characters, and revealing emotion. Considering the instructional purposes of the dialogue as a priority in our educational VR experiences and the development of a story arc, we defined two major functions for the use of dialogue in our VR storytelling experience design. First, dialogue is used to initiate actions and move the story forward. Meaningful dialogue often leads to specific character actions and keeps audiences engaged [17]. Second, dialogue can be used to convey learning information. With a multiple character setup, we rely on the dialogue as a vehicle to convey dry learning concepts in a fun and easy-to-understand manner through interactions between virtual agents.

Additionally, we utilized dialogue in following purposes to strengthen the storytelling experience:

- Express characters’ personalities
- Express emotions
- Distinguish characters through the way they talk

In our implementation, in higher levels of storytelling richness in which there are three or more characters (Figure 5), we introduced each new learning concept or topic via character dialogue and followed it by either directing the user to some visual and animations, or giving the user some interactive tasks related to the concept mentioned in the dialogue.

3.4.5 Voice Acting. Rich character and dialogue designs in our storytelling prototypes need voice acting to bring them alive. Therefore, a crucial part of our experience design is to obtain recordings of actors playing the roles of all speaking characters. To start with, we made some decisions on the gender of voices for specific characters. Since the characters PHIL, ARA-C, MAC, and MONO6 are those who speak a lot, we decided to choose female voices for them due to the fact that people generally respond positively to female voices in virtual agents and assistants [18]. We chose a male voice for the character MAJOR16 for its immune system commander role and another male voice for the MUTANT for its vicious nature as a villain in the story. We then recruited voice actors from Cinema major students and chose the people that met our selection criteria:

- The voice actor’s audition performance should match what we envisioned a character’s voice to sound like based on its role in the story. For example, the main character PHIL should have a young female’s voice with distinct and strong emotions, and MAJOR 16 from the command center should have an authoritative voice.
- The voice actor’s voice should have some flexibility to accommodate necessary changes we requested for the character role for which he/she is cast.

The whole recording process was conducted in an adaptive and actor-friendly way, including preliminary script reading, rehearsal with voice actors, script adjustment with feedback from voice actors, and Zoom recording sessions.

3.5 Experiment Design

3.5.1 Levels of Storytelling Richness. Our goal was to create three different storytelling designs within an immersive and interactive VR learning experience that lasts 20-25 minutes. Defining different levels of storytelling richness in each experimental condition presented challenges to us. On the one hand, there is no existing literature and applications on the levels of storytelling richness we can refer to. On the other hand, the complexity of a story and the richness of storytelling elements in it are greatly restrained by the length of the story itself. For example, a story that is told through a TV series across several seasons will have much more detail than its theater counterpart that normally runs within 2 hours.

Our design strategy to achieve different levels of richness in a short story was to focus primarily on the number of unique characters and the detail and drama of their dialogue. Secondarily, we also tweaked the richness of other storytelling elements such as conflict, resolution, and point of view. Specifically, at the minimal level of storytelling richness, we created only one speaking character as a virtual guide to convey learning concepts to the users through voiceover in 2nd-person and 3rd-person narrative voices. The conflict and resolution were voiced by a virtual guide as simple missions and tasks that are commonly used in a gameplay setting. The total experience is very similar to those seen in many educational VR applications. By contrast, at an advanced level of storytelling richness, we created a rich set of speaking characters with engaging dialogue narrated mainly through a 1st-person voice, and the conflict and resolution were conveyed in a way that is personal.
Table 2: Comparison of level of richness setups in three storytelling conditions (see supplementary content for full story scripts)

<table>
<thead>
<tr>
<th>STORYTELLING ELEMENTS</th>
<th>LEVEL I MINIMAL STORYTELLING</th>
<th>LEVEL II BASIC STORYTELLING</th>
<th>LEVEL III ADVANCED STORYTELLING</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHARACTERS</td>
<td>Total number: 1</td>
<td>Total number: 3</td>
<td>Total number: 6</td>
</tr>
<tr>
<td></td>
<td>Speaking non-player character (virtual guide)</td>
<td>Protagonist: non-speaking (user embodies)</td>
<td>Protagonist: speaking (user embodies)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Antagonist: non-speaking</td>
<td>Antagonist: speaking</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Supporting characters:</td>
<td>Supporting characters: speaking</td>
</tr>
<tr>
<td>DIALOGUE</td>
<td>None</td>
<td>Between supporting characters</td>
<td>Between protagonist and supporting characters</td>
</tr>
<tr>
<td></td>
<td>Convey learning information through simple voiceover</td>
<td>Convey learning information through dialogue among speaking characters</td>
<td>Convey learning information through rich dialogue among speaking characters</td>
</tr>
<tr>
<td></td>
<td>Narrated in 2nd and 3rd person voices</td>
<td>Narrated mainly in 2nd person voice</td>
<td>Narrated mainly in 1st person voice</td>
</tr>
<tr>
<td>CONFLICT &amp; RESOLUTION</td>
<td>Conveyed through simple missions and tasks voiced by virtual guide</td>
<td>Conveyed through dramatic conflict voiced by secondary characters</td>
<td>Conveyed through dramatic conflict that is personal to and voiced by the main character (embodied by the user)</td>
</tr>
</tbody>
</table>

One thing to be noted is that we decided to use the same plot structure in all three conditions as a method of experimental control, so that the users in the three conditions would have the same amount of exposure to the embedded learning concepts.

The study used a between-subjects design with level of storytelling as the independent variable. Each participant was randomly assigned to one of the three conditions: minimal storytelling (MS), basic storytelling (BS), or advanced storytelling (AS).

### 3.5.2 Integration of Interactivity into Linear Storytelling

Our previous study on level of interactivity on VR learning [19] suggested that higher levels of interactivity in an educational VR experience contributes to higher learning engagement and more experience enjoyment. While storytelling is the main component of the experience to explain complex immunology concepts to the students, adding rich interactivity to the experience allows them to apply what they have learned to various tasks and reinforces their understanding of the concepts. We experimented with a guided interactive narrative approach proposed by Weißand Müller [20] to prioritize the plot and the overall linear path of the story, but at the same time give the learners superposed interactivity at localized scenes (Figure 6).

Specifically, the interactivity we integrated into the story combines system-controlled virtual travel and full user-controlled task performance and localized virtual movement. We chose such a medium level of interactivity for the experience prototype because it allowed us to confine the user’s travel in the VE to a predesigned route with system-automation in order for them to experience intended story events, but at the same time still give them freedom to interact with virtual agents through learning activities.

### 3.5.3 Measures. Objective measure

We used a pre-test to measure participants’ prior knowledge of the learning concepts, an immediate post-test right after the VR experience to measure participants’ learning gains, and a delayed post-test 7 days later to measure study participants’ retention of the learning concepts. Each of the three tests had the same 17 items in the form of multiple-choice questions. Participants were not given any feedback or the correct answers
after taking the pre-test and post-tests. The order of the choices was shuffled randomly in the post-tests in order to reduce the possibility of memorization of the choices. The tests were scored on a 17-point scale, with each item being worth one point.

Subjective measures. We used three questionnaires (see supplemental content for details) and a short interview to collect participants’ feedback on learning and user experiences. Specifically, we developed a learning perception questionnaire (LPQ) 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; and effectiveness of the audio and storytelling elements on concept learning.

To measure the participants’ engagement in the storytelling design in the experiences, we adopted and modified a narrative engagement questionnaire (NEQ) that was proposed by Busselle and Bilandzic [21].

We also included the User Experience Questionnaire (UEQ) [22] to measure both classical usability aspects (efficiency, perspicuity, dependability) and user experience aspects (originality, stimulation) of our interactive VR application.

To get more insights about participants’ learning experiences, we conducted a short interview (see supplementary content for details) with each of them after their VR learning experiences. We used thematic analysis (TA) to analyze audio transcripts from interviews for each participant and tried to identify strong patterns or themes.

3.5.4 Apparatus. The VR experience was designed in Unity 2019 and was sideloaded in the Oculus Quest 2 stand-alone VR system, which includes a high-resolution headset, two six-degree-of-freedom controllers, and built-in tracking.

3.5.5 Participants and procedure. A total of 64 participants (32 females, 32 males, aged 19-70, mean age 27 years old) were recruited for the study from 24 different academic programs at a large research university. Participants were provided informed consent and completed the pre-test online at least 24 hours before their scheduled study session. They were particularly instructed not to look for answers from Google when doing the pre-test if they didn’t know answers to the questions. During the study session, each participant was randomly assigned to one of the three storytelling conditions and given the corresponding experience prototype to play. We assigned 21 participants to the minimal storytelling group (MS), 21 to the basic storytelling group (BS), and 22 to the advanced storytelling group (AS). Immediately after the experience, the participant was given the post-test and three questionnaires to fill out. After that, we asked them nine questions through a short interview and recorded their answers. One week later, the participants were given a delayed post-test to conclude the whole study.

3.6 Results

3.6.1 Objective Learning. We used the SPSS Statistics software package to analyze all quantitative data (quiz scores) from the study. We first did several calculations between participants’ pre-test and immediate and delayed post-test scores using one-way ANOVA. Specifically, we calculated:

- Immediate learning gains: difference between pre-test and immediate post-test scores
- Delayed learning gains: difference between pre-test and delayed post-test scores
- Learning retention: difference between immediate and delayed post-tests

We then did a paired sample T-test on quiz scores in all three storytelling conditions together and found a significant difference between pre-test scores (M=7.52, SD=2.74) and immediate post-test scores (M=12.92, SD=2.85); (t(63)=-14.00, p=0.001), and pre-test scores (M=7.52, SD=2.74) and delayed post-test scores (M=12.59, SD=2.56); (t(63)=-14.11, p=0.001), which indicates that no matter the storytelling condition and post-test time, participants did learn the information to a large extent after the experience (Figure 7).

We ran a one-way between-subjects ANOVA to compare learning gains and retention across the three different conditions. We
were not able to find any significant difference between immediate and delayed post-test scores across the three storytelling conditions. Participants in all three conditions achieved similar immediate learning gains, delayed learning gains, and learning retention. In addition, there was not a significant loss of knowledge one week after the experience, considering the lack of difference between the immediate and delayed post-test scores.

### 3.6.2 Quantitative Questionnaire Data

Kruskal-Wallis tests with one-way ANOVA were used to analyze non-parametric data from the three questionnaires (LPQ, NEQ, and UEQ). Our analysis of LPQ data returned no significant differences in any of the statements between different storytelling conditions.

The NEQ contains 12 statements related to a participant’s perceptions on narratives in the experiences. Our analysis of NEQ data returned no significant differences in any of the statements between different storytelling conditions.

The UEQ measures six attributes of a product experience (attractiveness, perspicuity, efficiency, dependability, stimulation, novelty). Our analysis revealed no significant effects on user experience between different storytelling conditions. However, all three conditions were rated in the ‘good’ or ‘excellent’ categories in all six aspects of user experience, with only one exception—the basic storytelling condition rated just below ‘good’ in the category of perspicuity.

### 3.6.3 Qualitative Interview Data

We asked each participant 9 questions related to the condition they experienced to collect more information about their VR learning experience with different levels of storytelling richness. We recorded participants’ answers to the questions and transcribed the audio. We then used the method of thematic analysis [23] to analyze all participants’ transcripts. Possible themes were identified from the transcripts after two rounds of coding. Questions related to our hypotheses and their coding results are shown below.

Q1 asked about the participants’ general experience of learning immunology concepts in VR. We found their answers centered around two themes: hedonic factors of using VR, such as coolness, enjoyment, and fun, and various aspects of the VR experience that they perceived to facilitate learning, including arousing interest and increasing engagement, visualization of concepts, storytelling, interactivity, embodiment, gaming elements, and immersion (see Table 3).

For hedonic factors, we found that participants at the higher levels of storytelling richness reported more enjoyment, fun, coolness, and interest (8 in AS group and 7 in BS group, versus 4 in MS group). In terms of specific factors facilitating learning, there were common factors reported by participants from all three groups that contributed to their learning, such as visualization, interactivity, and immersion. However, more participants in the basic and advanced storytelling groups perceived that storytelling (6 in AS and 4 in BS versus 2 in MS) and engagement (4 in AS versus 1 in MS) were factors contributing to their learning.

Q2 was intended to collect information on factors that positively contribute to learning in VR experiences based on the participants’ own perceptions. We found several major themes shown in Table 4. Participants thought that factors such as interactivity, gameplay elements, storytelling, character dialogue, visualization, embodiment, and immersion, contributed to their learning most. There were a few more participants in basic and advanced storytelling groups (6 in BS and 5 in AS versus 4 in MS) who mentioned storytelling as an important factor contributing to their learning. Additionally, many more participants in BS and AS groups than in the MS group (6 in BS and 5 in AS versus 1 in MS) mentioned character dialogue/voices as an important factor contributing to their learning.

![Figure 7: Boxplots of the results across three conditions.](image-url)
Table 3: Emerging themes from Q1

<table>
<thead>
<tr>
<th>Themes</th>
<th>Participants addressing the theme</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimal Storytelling</td>
</tr>
<tr>
<td>Hedonic factors</td>
<td>4</td>
</tr>
<tr>
<td>Facilitating learning (in various ways, see sub-themes below)</td>
<td>3</td>
</tr>
<tr>
<td>Non-specific</td>
<td></td>
</tr>
<tr>
<td>Arouse interest and motivation</td>
<td>0</td>
</tr>
<tr>
<td>Engagement</td>
<td>1</td>
</tr>
<tr>
<td>Storytelling</td>
<td>2</td>
</tr>
<tr>
<td>Visualization</td>
<td>8</td>
</tr>
<tr>
<td>Interactivity</td>
<td>2</td>
</tr>
<tr>
<td>Embodiment</td>
<td>3</td>
</tr>
<tr>
<td>Gaming elements</td>
<td>1</td>
</tr>
<tr>
<td>Immersion</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 4: Emerging themes from Q2

<table>
<thead>
<tr>
<th>Themes</th>
<th>Minimal Storytelling</th>
<th>Basic Storytelling</th>
<th>Advanced Storytelling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactivity</td>
<td>7</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Gameplay elements</td>
<td>9</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>Storytelling</td>
<td>4</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Character dialogues/voices</td>
<td>1</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Visualization</td>
<td>9</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Embodiment</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Immersion</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5: Emerging themes from Q5 and Q6

<table>
<thead>
<tr>
<th>Themes</th>
<th>Minimal Storytelling</th>
<th>Basic Storytelling</th>
<th>Advanced Storytelling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distinct characters</td>
<td>1</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Story</td>
<td>6</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Character dialogues and voices</td>
<td>1</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Emotions</td>
<td>1</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Empathy</td>
<td>0</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Embodiment</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Q5 and Q6 asked the participants their perceptions of storytelling elements when playing the VR learning experiences. The data showed that in both the basic and advanced storytelling groups, there were more participants who agreed that the experiences they played had a strong story (17/21 in BS and 15/22 in AS versus 11/21 in MS) and strong characters (15/21 in BS and 17/22 in AS versus 9/21 in MS). When asked why they thought so, more participants in basic and advanced storytelling groups mentioned reasons such as distinct characters, plot, character dialogue and voices, emotions, empathy, and embodiment, as shown in Table 5.

Q9 asked the participants their perceptions of the effectiveness of learning immunology through VR storytelling experiences. Our data showed that most participants in all three groups agreed that learning immunology through VR storytelling was effective, although the percentage of agreement was slightly higher in the basic and advanced storytelling conditions (16/21 in MS, 19/21 in BS, and 21/22 in AS). As for the reasons, they centered around the following themes: engagement, visualizations of the learning concepts, gameplay elements, the story, interactivity, and help with retention of learning concepts. We found that more participants in the basic storytelling and advanced storytelling groups mentioned that the embedded story helped with their learning during the VR experience, and they retained more learning information and felt more confident when taking the post-tests (Table 6).

3.7 Discussion

Recall that our hypotheses for the study were:

- H1. Level of storytelling richness will significantly affect a learner’s learning gains.
- H2. Stories with higher levels of richness will engage learners more in an educational VR experience.
H3. The learners in the advanced level of storytelling richness group will have the highest rate of retention of learning information one week after the VR experience.

H4. Compared to participants in the lower level of storytelling richness condition, more participants in higher level of storytelling conditions will perceive storytelling elements as useful factors for effective learning.

Regarding H1, our quantitative data from the pre-test and two post-tests did not support it, because there were no significant differences in immediate learning gains or delayed learning gains among the three storytelling conditions. Participants from all three conditions achieved uniformly high learning gains and learning retention. Subjective ratings from post-experience questionnaires on participants’ learning experiences also showed no significant differences, which indicates that all three storytelling conditions were effective for immunology learning. We speculate that there may be different factors in each condition that contribute to a similar effectiveness. Although the interactivity and gameplay elements were the same in all three conditions as a way of experiment control, they may have different contributions to learning when combined with different levels of storytelling richness. For example, the minimal storytelling condition had the least amount of storytelling richness but was effective, which may be due to an appropriate level of interactivity and gameplay elements that led to engagement and memorability for some of the key learning concepts, especially since the neutrophil’s killing mechanisms were shown primarily through interactivity. The advanced level of storytelling had the most storytelling richness and was effective for learning. It is possible that the richer storytelling elements made it more compelling and memorable and played a dominant role in contributing to the effectiveness of learning.

Regarding H2, statement 2 from the LPQ questionnaire specifically addresses the issue of engagement. However, we found that participants’ ratings of the statement were also uniformly high (4.5–4.8 out of 5) in all three conditions, meaning that all levels of storytelling richness were engaging, and there wasn’t any significant difference that could be found between them statistically. Therefore, this hypothesis was not supported. One possible explanation may be that the interactivity was at the same level in all three conditions, and interactivity and gameplay elements are often positively associated with engagement [19, 24]. It is possible that the effects of interactivity and gameplay elements were already strong enough that they contributed to the engagement and created a ceiling effect, which made the contribution of storytelling richness hard to measure using our current experimental design. It is also possible that both interactivity and storytelling contributed differently to the engagement in different storytelling conditions, but the combined effect of both resulted in a similar amount of engagement in each condition. One thing to be noted is that although the hypothesis was not supported by our survey data statistically, our interview data from Q1 showed that there were more participants in the advanced storytelling group than in the minimal storytelling group (5 in AS versus 1 in MS) who mentioned engagement as a positive factor facilitating learning in their VR experiences. This finding may indirectly suggest that higher storytelling richness from the advanced storytelling condition did lead to engagement in learning for some participants.

To address H3, we looked into the difference between the delayed post test and immediate post test as a measurement of knowledge retention for a period of 7 days. However, there was no evidence of significant difference statistically to support this hypothesis. Participants from all three storytelling conditions had a similar amount of knowledge retention. Apart from the possible explanations already offered for hypotheses 1 and 2, another possible reason for this finding is that the interval between the two post-tests may be too short, so that most participants still have a good memory of what they learned in the VR experiences. To measure long-term memory of knowledge, some educational researchers suggest using longitudinal measures across a wide time frame of weeks to several months to even years [25].

The H4 can be evaluated from study participants’ interview data. We didn’t find strong evidence to support this hypothesis. However, some differences in participants’ response counts from interview questions Q1, Q2 and Q9 are worth-mentioning. Specifically, Q1 asked the participants their general feeling of VR learning experiences. We found there were more participants in the advanced and basic storytelling groups who mentioned the theme “storytelling” as a factor that facilitated their learning in VR (6 in AS and 4 in BS versus 2 in MS). Q2 asked the participants what design factors they thought were the most useful during their VR learning experiences. More participants in the higher levels of storytelling groups mentioned character dialogue, which is one essential element of storytelling, as a useful design factor for a positive learning VR experience (5 in AS and 6 in BS versus 1 in MS). Q9 asked the participants if they thought the experience, overall, was an effective way to learn about immunology, and why. More participants in the higher levels of storytelling groups agreed the VR experience was effective for learning (21/22 in AS, 19/21 in BS, versus 16/21 in MS). When asked for the reasons, there were more participants in the advanced and basic storytelling groups (5 in AS, 3 in BS versus 1 in

### Table 6: Emerging themes from Q9

<table>
<thead>
<tr>
<th>Themes</th>
<th>Minimal Storytelling</th>
<th>Basic Storytelling</th>
<th>Advanced Storytelling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engagement</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Visualization of learning concepts</td>
<td>7</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Gameplay elements useful</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Embedded story</td>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Interactivity</td>
<td>6</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Help retain learning information</td>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

We can explore these themes further to understand their implications for educational design in VR environments.
MS) who mentioned that it was the embedded story in the experiences that helped them retain a good amount of information and gave them confidence in the post-tests. However, due to the small sample size in the study, these small differences are not enough for us to make a conclusion that this hypothesis is fully supported by the interview data.

In summary, we didn’t find evidence to support the major study hypothesis that level of storytelling richness significantly affects a user’s learning gains and retention, nor did we find evidence to support the hypothesis that higher level of storytelling richness will lead to higher learner engagement. There are several reasons that may account for these findings. First, participants in all three conditions achieved uniformly high knowledge gains and retention, which indicates that all storytelling conditions were quite effective for immunology learning. We speculate this is because: a) they all had the same amount of interactivity and gameplay elements, and the participants generally enjoyed them during learning; b) they all included the same story events to embed the same amount of immunology concepts and learning activities; c) they all fully immersed the participants in a compelling and novel environment; and d) they all helped the participants focus on the learning content during the experience play. These design factors, although being carefully controlled in all three conditions, may have had a greater combined contribution to learning and engagement than that solely from the storytelling elements, thus creating a ceiling effect. The similar uniformly high ratings in both the LPQ and UEQ also suggest a possible ceiling effect.

Second, it is a general challenge in media comparison studies that it is very difficult to find actual learning differences as measured by pre- and post-tests, due to individual differences and many other uncontrollable factors outside the materials being studied [26, 27]. For example, novelty may be a factor that affected learning perceptions of first-time VR users. We observed that novice VR users were generally very excited and engaged during the VR learning experiences no matter which experiment group they were in.

Third, the medium level of interactivity in all three conditions may have a greater effect on learning than that of the storytelling elements in the VR experiences. We chose the medium level of interactivity to be integrated into the study prototype because it worked well with our proposed learning model of interactive VR storytelling.

Our interview data revealed some other findings, unrelated to the hypotheses, that are worth discussing here. First, more participants in the advanced storytelling group associated the themes “Emotions” and “Empathy” with a strong story and strong characters in the experience, compared to many fewer people in the basic storytelling group and very few people in the minimal storytelling group (emotions: 9 in AS, 2 in BS, 1 in MS; empathy: 5 in AS, 2 in BS, 0 in MS). We believe this is due to the additional richness of storytelling elements in characters and dialogue. Since empathy can provoke a better attitude towards learning when there is interaction between learners and virtual agents [28] and increase engagement with interactive agents [29], we hypothesize that both emotion and empathy triggered by virtual character interaction from the advanced level of storytelling may be positive factors that affect a user’s learning experience.

### 3.8 Limitations

Two limitations of this study can be addressed for future improvements. First, although we carefully controlled design elements other than the storytelling elements in the three conditions to eliminate possible confounding factors, we didn’t fully consider the possibility that some design elements like interactivity and gameplay may have a greater impact on learning than the storytelling elements, making the measurement of the latter difficult. Future experiments need to further consider the individual effects of design elements on learning and remove design elements that have competing effects on learning with the storytelling elements.

Second, we left a short interval of 7 days between immediate and delayed post-tests to measure participants’ learning retention and didn’t find any significant differences among the conditions. A longer interval of several weeks or months may be necessary in order to measure differences in long-term memory of knowledge gained.

### 3.9 Conclusion and Future Work

This study investigated how different levels of storytelling richness in interactive and immersive VR experiences affect learning outcomes and user experiences. Although there was no evidence of significant differences found between different experiment conditions from objective and subjective quantitative measures, participants in higher levels of storytelling groups had a higher rate of perceiving storytelling elements as the most useful features that helped their learning and knowledge retention in the VR experiences. Additionally, more participants in the advanced storytelling group reported that they felt empathy and emotions from the embodiment of the main character, which can be a positive sign that they were psychologically more connected with the virtual character and the story world. This tentative finding should be further studied in the design of interactive and immersive VR storytelling experiences for learning.

Based on our empirical data and findings, we suggest the following design guidelines for the storytelling design in interactive educational VR experiences:

- When integrating interactivity with rich stories, adjust its intensity to avoid cognitive overload.
- Utilize emotion and empathy elements to promote learning in rich stories.
- Select suitable learning concepts for interactive VR storytelling experiences.

Our study findings suggest two possible directions for our future work. On the one hand, we could further investigate the level of storytelling richness on learning with some revisions to the current study: a) we could minimize the interactivity in future experiment designs by using system automation to replace all user-controlled interactivity. b) we should develop measures that allow us to measure the individual effects of different design elements on learning. On the other hand, since more participants in the advanced level of storytelling group in our current study reported emotional and empathetic feelings with the main character they embodied, we are curious how empathy that is aroused from the higher levels of storytelling richness may contribute to positive learning experiences in VR. Unfortunately, we were not able to measure it directly in the
current experiment, because it was a factor that only emerged in post-hoc analysis. Therefore, it is necessary to design a future study that specifically measures the effect of empathy from advanced levels of storytelling richness on learning in an educational VR experience.

REFERENCES
[22] Martin Schrepp. 2015. User experience questionnaire handbook: All you need to know to apply the UEQ successfully in your projects