Measuring Narrative Engagement in Interactive Cinematic VR Experiences

Anonymized[1], Anonymized[1], and Anonymized [1]

Abstract. This research involves the development of a cinematic VR experience that exhibits narrative engagement and the investigation of possible measurement tools to evaluate that engagement. This is accomplished by the implementation and analysis of standardized self-reporting measures and observational data. The efficacy of these measurement tools is discussed as well as their possible modifications and limitations for storytelling in VR.

Keywords: Immersive Storytelling, Interactive Storytelling, Cinematic VR

1 Introduction

This research investigates a range of measurement tools towards assessing narrative engagement in interactive cinematic VR experiences. This article describes the design of a cinematic VR experience and its evaluation using standardised self-reporting measures.

For the context of this research, cinematic VR describes immersive experiences with limited interactivity and a strong emphasis on storytelling. Cinematic VR productions are not game-like experiences, but VR narratives based on targeted design and psychological criteria supported by the technology that VR inherently provides.

This research is important as VR is becoming more prevalent in research and personal use. However, the medium still has an untapped potential for immersive storytelling. Additionally, narrative engagement within the VR storytelling experience is both difficult to create and to evaluate, thus it is valuable to investigate work dedicated to VR storytelling as opposed to a game like experience. Lastly, this work proposes to

develop a dedicated high-quality experience expressly for the purpose of investigating narrative engagement.

2 Research Design

2.1 Project Creation

To explore how narrative engagement could be measured in a VR experience, we designed a VR experience focused on narrative engagement and storytelling. The VR experience created for this study was made based on the recommendations and findings from [1]. This creation comprised three main phases: the Script, the Assets, and Interactivity and immersion.

The Script: Traditional western story tropes might be ill-suited to VR, due to the possibility that the model of the story line would break because of the immersive and interactive nature of VR [2]. With this in mind, a more general outline was chosen for the script. Using a simplified version of Blake Snyder's [2] Beat Sheet as a guide, an initial script was sketched out and was loosely based on an accumulation of varied Scottish folklore books for content.

The concept of change [3] was woven into the script early on to assist with engaging the user from the beginning during the catalyst and debate. It was again employed towards the end, in all is lost, as a final turning point. As Richardson et al. [4] postulated that listening was an active process in co-creation, the script was written to be narrated, filling in details that were not present in the world, as well as leaving out details that were.

The Assets: The assets for the project encompassed the concepts of characters, story-world, and curiosity. For the character creation, the characters were given a personality based on the FFM [5]. On this scale, the main character was given high scores for openness, conscientiousness, and agreeableness, with low scores for extroversion and neuroticism. This meant that the character's personality was curious, dependable, reserved, empathetic, and calm. This created a blueprint for how the character would look and act and was able to be designed based on these personality traits. To emphasise the personality and identity of the character, other assets were created as behavioural residue [6]. An example of such assets were items like a smoking pipe, picture frames, maps, and wine bottles that gave small indications about the character's life, many of which were interactive. To solidify the character, the story-world was then created based on the character's attributes and persona. The world itself was created on an island, with the scenes occurring in various locations around it; this was done so that when scene changes occurred, it would lessen the amount of time it took for the user to reorient themselves in the world, since they could see all the other places they had previously been. To bring the story-world to life, life was added through other assets such as birds, rabbits, sea creatures, grasses, and trees.

Fig. 1. Example scene from project

Some of these played multi-purposed roles, contributing not only to the story world and persona of the character but were also employed as curiosity types [7]and diegetic devices [8, 9]. An example of this is a recurring bird whom the user first meets in the menu, and then again in the first scene where it can be interacted with. It is then placed throughout various other scenes to help direct the focus and attention of the player (See Figure 2).

Fig. 2. Example map diegetic placement

Curiosity played a dual role in both the assets and the interactivity. Curiosity types [7] are categorized in digital narratives as: manipulatory, complex/ambiguous, perceptual, conceptual, and adjustive/reactive. For the creation of the experience, 4 out of 5 types were employed. Manipulatory was introduced simply by the use of the controller in the experience, with the ability to grab, hold, or throw items. Complex/ambiguous was employed by making complex objects to interact with. Some of these included objects such as birds or rabbits that were animated and provided haptic feedback when touched. Others were in the form of picture frames that highlighted or changed their image when handled. Perceptual was implemented through music, sound cues from various objects, and visual highlights. And adjustive reactive constituted the items that were simpler and had a common use, such as a violin that the user could play. Conceptual was left out, simply due to the difficulty of executing it within the narrative.

Interactivity and immersion: As stated above curiosity also played a role in interactivity, as many of the "curious" objects were also interactive. Along with this, the use of music, audio cues, and highlights were used to focus/gain the user's attention and increase immersion. In particular, both ambient and spatial sound were used throughout the VR experience [10]. Music and narration were ambient with no discernable source. The spatial sound encompassed everything else. This included elements like waves crashing, bird calls, wind, thunder, rain, and whale calls. Each sound had an individual attenuation radius (the falloff of the source) utilising a natural sound function (See Figure 4) and employed binaural spatialisation (the sound changed and shifted based on the user's physical orientation towards the sound). These overlapped with each other to create a more natural environment.

Fig. 3. Example map of spatial sounds

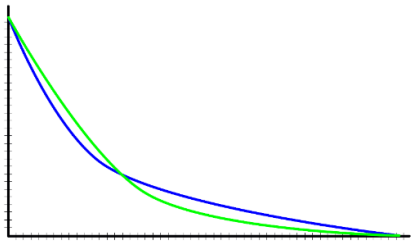

Fig. 4. Natural sound function that determines the rate of attenuation over distance. This models a naturalistic falloff behaviour that is closer to matching reality.

In addition to the spatial and ambient sounds, other interactivity was built into various other assets. As mentioned earlier with regards to curiosity, some picture frames changed and were highlighted when held, other objects sound be collected and thrown or placed down by the user such as vegetables and wine bottles.

2.2 Evaluation Protocol

This section describes the self-reporting measures and the observational methods employed in this study along with a discussion on ethical considerations for these two methods.

Self-Reporting Measures

As narrative engagement is a multifaceted concept, several questionnaires can be employed for a well-rounded scope in the context of interactive storytelling VR experiences. For this research, the questionnaires used were based on the following concepts:

- Narrative understanding
- Attention Focus
- Narrative presence
- Emotional engagement
- Suspense
- Curiosity
- Flow
- Presence
- Enjoyment
- Aesthetic pleasantness

For consistency, all scales were measured on a 5-point Likert scale using a combination of forward scoring **(F)** and backwards scoring **(B)**, as denoted on the following scales. Forward scoring has numerical values attached to the anchors in a forward direction, with fully agree = 5, and fully disagree = 1. Backwards (reverse) scoring has numerical values attached to anchors in the opposite direction, with fully disagree = 5 and fully agree =1.

Narrative Engagement Scale

Busselle and Bilandzic's [11] research interpreted four factors for narrative engagement while developing their Narrative Engagement Scale (NES). These were narrative understanding, attentional focus, emotional engagement, and narrative presence. Although it was not developed specifically for VR, it has been a widely used model in research and other VR studies [12, 5, 13] as well as having a Cronbach's α of over .80. For these reasons, it can potentially be adapted for use in interactive VR stories. The NES consists of 12 questions on a 7 -point Likert scale and was adapted to a 5-point scale for consistency across all questionnaires.

Narrative understanding

- At points, I had a hard time making sense of what was going on in the experience. (B)
- My understanding of the characters is unclear. (B)
- I had a hard time recognising the thread of the story. (B)

Attentional focus

- I found my mind wandering while the during the story experience. ***(B)***
- While in the virtual world I found myself thinking about other things. ***(B)***
- I had a hard time keeping my mind on the story. ***(B)***

Narrative presence

- During the experience, my body was in the room, but my mind was inside the world created by the story. **(F)**
- The experience created a new world, and then that world suddenly disappeared when the application ended. **(F)**
- At times during the experience, the story world was closer to me than the real world. **(F)**

Emotional engagement

- The story affected me emotionally. **(F)**
- During the experience, when a main character succeeded, I felt happy, and when they suffered in some way, I felt sad. **(F)**
- I felt sympathy for some of the characters in the story. **(F)**

Suspense Scale

Measuring suspense in interactive storytelling is a somewhat novel idea. Knobloch et al [14] developed a three-item scale for suspense rating media content in terms of being thrilling, gripping, and exciting. Other scales used to measure suspense are context specific [15], but neither of these are in the context of interactive narratives. Based on these studies and his own research, Roth [16] postulated that the measurement of suspense of interactive narratives should be based on the emotional involvement in the story's outcome. He therefore constructed 10 items to capture suspense based on emotional investment in the story specifically in the context of interactive narratives. This scale was later shorted to four items, based on the items with the highest item-total correlations.

- what would happen next **(F)**
- Sometimes I was worried about how the story would develop. **(F)**
- Some moments were rather suspenseful. **(F)**
- I found myself wishing for a particular story outcome. **(F)**

Curiosity Scale

.

Spielberger et al. [17] determined curiosity as a state, thus the State-Trait Curiosity Inventory (STCI) was developed to measure the intensity of curiosity as a transitory emotional state (17, 18]. The STCI includes 10 items on a 4-point scale asking participants to report how they feel at a particular moment. This was adapted to a 5-point Likert for consistency throughout the other questionnaires, and "in the moment" was rephrased to "during the experience". Additionally, the 10 items were adapted into three based on the recommendations [16].

During the experience I felt…..

- Curious (F)
- Interested (F)
- Inquisitive (F)

Flow Scale

Csikszentmihalyi [19] proposed eight factors for optimal flow: challenge activity; merging of acting and awareness; clear goals; direct immediate feedback; concentration; a sense of control; loss of self-consciousness; and an altered sense of time. Based on this model, the Flow State Scale (FSS) was developed [20]. Initially, this scale was a 36-item list, and later paired it down to 9 items to allow for usage in a wider range of studies. Each item chosen reflected one of the nine higher order factors from the original scale [21]. Findings from the shorter list revealed that it provided a good representation of the long version with high reliability. This was adapted the scale into five items based on the highest item-total correlations.

During the experience. . .

- . . . I felt competent enough to meet the demands of the situation (F)
- . . . I acted spontaneously and automatically without having to think (F)
- . . . I had a strong sense of what I wanted to do (F)
- . . . I had a good idea while I was performing about how well I was doing (F)
- . . . I was completely focused on the task at hand (F)

Presence Scale

There are currently a few standardised presence questionnaires in circulation for VR applications [22, 23, 24]. The IPQ [23] was chosen based on research of the efficacy of presence scales [25] as it provided the highest reliability within a reasonable timeframe. The IPQ is a 14-item list, on a 5-point Likert scale. The items consist of 4 categories: General, Spatial presence (the sense of being physically present in VR), Involvement (measuring the attention devoted to the experience) and Experienced Realism (measuring the subjective experience of realism. Based on these categories, the scale was shorted to contain one item from each category.

- In the experience I had a sense of "being there" (G) **(F)**
- I felt present in the virtual space (SP) **(F)**
- The virtual world seemed more realistic than the real world (ER) **(F)**
- I was not aware of my real environment (INV) **(F)**

Enjoyment Scale

The measuring of enjoyment has proved somewhat problematic. While the concepts of enjoyment have been used in media research (such as amusement, sense of achievement etc. [26], there is no study available that has attempted to measure it directly [16, 27]. Therefore, a simple short scale consisting of two questions was created.

The experience. . .

- . . . was entertaining (F)
- . . . was enjoyable (F)

Aesthetic Pleasantness Scale

Aesthetic pleasantness in media is often related to the visuals and audio. Aesthetic evaluations may relate to the physical appearance of characters or landscape imagery. Additionally, aesthetic content can relate to the personal background and previous experiences of the recipient. For instance, the depiction of a scene in a movie, can remind the viewer of feelings that resonate with the recipient's mood, thus evoking congruent feelings [28]. Thus, in the context, it is applied to encompass the elements of story-world, characters and emotion. For this study, the following questionnaire was used to access aesthetic pleasantness [29].

The experience. . .

- . . . made me think (F)
- . . . made me think about my personal situation (F)
- . . . told me something about life (F)
- . . . was inspiring (F)
- . . . moved me like a piece of art (F)

Self-Reporting Limitations

Self-reporting measures to reflect on past experiences can be somewhat limited, as it can be hindered by such things as selective memory, mixing memories of other events or exaggeration. However, there is still validity in the use of these methods, as these limitations can be reduced. One such reduction, is the use of standardized questionnaires as they can be backed with research and a high Cronbach α, increasing their validity. Additionally, wording of the questions was kept to the specific standard to avoid confusion or vagueness, with the exception of changing to the phrase "during the experience" across all scales for consistency.

Observation

Observational data for the study was recorded during the experience by the researcher in a nonparticipant role. The data recorded is in a semi-structured format using predefined events. The participants were aware that they were being observed, and aware that the researcher would not participate in the experience. The participates were also able to provide open-ended comments after the completion of the post questionnaire. An observational protocol was created for use during the observations. This included the current scene, time, and a record of events (See Table 1)

Table 1. Example of Observation Protocol for Scene 2

Scene	Time	Description of Events
2	1:10	Interaction with bird
	1:25	Following gaze of character
	1:40	Interaction with character

Observation was carried out via online video (Zoom), with the participant sharing their PC screen. This allowed the researcher to view both the participant and their camera view during the experience.

Limitations of observational data can include the researcher being seen as intrusive. The interruption of the experience to conduct survey or interviews can lead to a disruption of the flow, and thus lead to disengagement [13]. To mitigate this, the obser-

vational data recorded was non-invasive; participants were not asked questions during the experience.

2.3 Data Treatment

After the project was completed, 10 participants were recruited to take part in the experiment. The data collection was then divided into 2 main phases: *observation* and *reflection*.

Observation phase. Once the participants had their headset on, they started the program and observations were made and recorded throughout their experience. This data was qualitative in nature.

Reflection phase. After the completion of the VR experience, participants were invited to complete a set of self-reporting questionnaires. The Narrative Engagement Scale (NES); the Suspense Scale (SS); Curiosity Scale (CS); Flow Scale (FSS-2); Presence scale (IPQ), Enjoyment Scale (ES), and Aesthetic Pleasantness (APS). This data was quantitative in nature.

After completion of the data collection, the data analysis began consisting of 3 phases: *Analyse Quantitative, Analyse Qualitative*, and *Mixed Methods*. First, the quantitative results were analysed in terms of statistical results. Second, the qualitative database was analysed by coding the data and collapsing the codes into broad themes. The final phase is the mixed methods analysation, which consists of integrating the two databases. The integration of this data uses a data transformation approach; after the qualitative data had been coded into themes, they were counted and grouped, to form quantitative measures. The following sections will discuss the results of the quantitative and qualitative data.

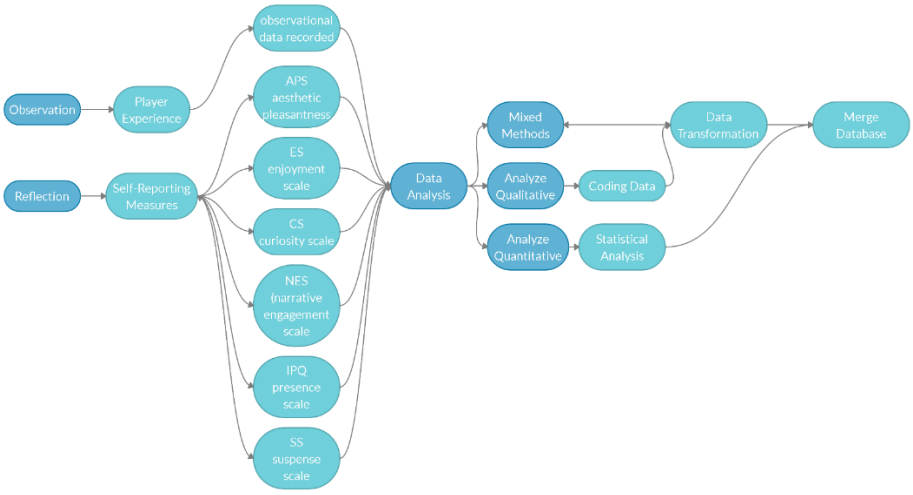

Fig. Error! No text of specified style in document.**-2.** Phases of Analysis

3 Results

3.1 Presentation of Quantitative Data

All data was measured on a 5-point Likert scale using a combination of forward and backwards scoring. In this section, the mean and standard deviation are provided for each of the scales. The full record of data is to be published at a later date.

The NES (Narrative Engagement Scale) [13] was considered by the individual factors:

- narrative understanding--the ease in comprehension of the story.
- attentional focus—concept that one should not be aware that one is distracted.
- emotional engagement--feeling for or with the characters.
- narrative presence—sensation that one has left the actual world and entered the story.

Narrative understanding recorded both the highest mean at 4.83, and lowest standard deviation at .37. Likewise, Emotional engagement and Narrative presence also recorded high values with low deviations. Attentional focus saw the lowest data with a mean of 3.60, and the highest variation at 1.57 (See Table 2).

Table 2. Narrative Engagement Scale Data

	Narrative understanding	Attentional Focus	Narrative presence	Emotional engagement
NES	M = 4.83	M = 3.60	M = 4.33	M = 4.50

(Narrative Engagement Scale)	SD = 0.37	SD = 1.57	SD = 0.82	SD = 0.76

Along with the individual factors, the NES was combined with the remaining scales for further analysis. Combined, the overall NES was recorded with a mean of 4.32 and the second lowest standard deviation of .45. The ES (Enjoyment Scale) recorded the highest at 4.75, and lowest deviation at .43. It is also important to note that the lowest scoring scales were the SS (Suspense Scale) at 3.35, with a fairly large deviation at 1.23, as well at the FFS-2 (Flow Scale) at 3.64. Although scoring fairly high comparatively, the IPQ (Presence scale) all showed a larger deviation at 1.25. These results and fluctuations in data will be discussed further during the analysis of this paper (See Table 3).

Table 3. Combined Assessment Data

NES (Narrative Engagement Scale)	M = 4.32	SD = 0.45
SS (Suspense scale)	M = 3.35	SD = 1.23
CS (Curiosity Scale)	M = 4.63	SD = 0.60
FFS-2 (Flow Scale)	M = 3.64	SD = 1.07
IPQ (Presence Scale)	M = 3.92	SD = 1.25
ES (Enjoyment Scale)	M = 4.75	SD = 0.43
APS (Aesthetic Pleasantness Scale)	M = 4.43	SD = 0.83

3.2 Presentation of Qualitative Data

Qualitative data for this study was gathered via observation during the VR experience. The data was hand recorded, then transcribed into documents and coded in to like themes. This data was coded into the following themes:

- Focus on Characters
- Interaction with Character
- Follow character gaze
- Interact with shell (*meaningful item*)
- Focus on crow (diegetic item)
- Aesthetic focus (*environment*)
- Aesthetic focus (*Life*)
- Interaction with other objects

Table 4. Codes at a Glance (Percentages)

Codes	Percentage of participants
Focus on Characters	100%
Interaction with Character	100%
Follow Character gaze	90%
Interact with Shell (Meaningful Item)	100%
Focus on Crow (Diegetic)	100%
Aesthetic focus (Environment)	100%

Aesthetic focus (Life)	100%
Interaction with other objects	80%

Table 5. Codes at a glance (Frequency per participant)

Code System	10	9	8	7	6	5	4	3	2	1	SUM
Interact with other objects	1	1	5	7	1	2		11	3		31
Aesthetic Focus (Life)	4	5	9	8	5	7	8	4	3	3	56
Aesthetic focus (Environment)	7	5	7	2	4	3	7	3	3	4	45
Focus on Crow (Diegetic)	2	1	1	3	1	3	3	1	2	2	19
Interact with Shell (Meaningfull Item)	2	2	2	1	2	1	1	2	4	1	18
Follow Character Gaze	2	1	6	2	1	1	2		2	1	18
Interact with Character	1	1	1	1	1	1	1	4	1	1	13
Focus on Characters	15	11	14	11	10	16	13	11	10	9	120

As Table 4 demonstrates, the majority of the codes appeared for each participant with the exception of follow character gaze, which was reported at 90 percent, and interaction with other objects which was reported at 80 percent.

It is important to note that while Table 5 reveals the frequencies of the codes per participant, and their totals, the totals are not necessarily an indication of priority of one code over another, as each code holds a different purpose. For this reason, each code will be evaluated and analysed independently.

Firstly, both the Focus on Crow (Diegetic) code and *Interact with shell* (meaningful) code had a max amount of 4 possible occurrences within the experience. The diegetic code revealed the following:

- 30 percent focused on the object 75 percent of the maximum allowance
- 30 percent focused on the object 50 percent of the maximum allowance
- 40 percent focused on the object 25 percent of the maximum allowance

Whereas the meaning code showed:

- 10 percent interacted with the object 100 percent of the maximum allowance
- 50 percent interacted with the object 50 percent of the maximum allowance
- 40 percent interacted with the object 25 percent of the maximum allowance

Conversely, *focus on characters, Aesthetic focus (life)* and *aesthetic focus(environment)* did not have a set number of occurrences. Therefore, their frequencies and totals are of some importance at 120, 56, and 45 respectively. *Interact with character* only occurs a total of 13 times, however it is important to note that all participants attempted interaction with the character at least once, and that single interaction occurred at the same point during the story. *Interact with other objects* is wildly varied as far as frequencies go and holds a larger SD of 3.37. The final code is the

follow character's gaze, which held an average of 2 per participant with a deviation of 1.8.

With both sets of data presented, the following section will explore a deeper analysis of each, followed by a comparative analysis of the combined results.

4 Analysis

4.1 Self Reporting Analysis

The Narrative Engagement Scale (*NES*) [13] which consisted of 4 separate subcategories:

- narrative understanding
- attentional focus
- emotional engagement
- narrative presence

The data from this scale demonstrated high averages and low deviations for all categories save for one, *attentional focus*. Although the mean was above average (3.60) the deviation was high (1.57). The following statements were used for this subcategory:

- I found my mind wandering while the during the story experience.
- While in the virtual world I found myself thinking about other things.
- I had a hard time keeping my mind on the story.

As demonstrated, these statements primarily focus on the mind: paying attention and not wandering from the subject. The discrepancy for this large deviation of the subcategory can be narrowed into two potential factors. The first possibility is the differences of cognitive capabilities and personalities of each participant. As no baseline was gathered before the experience on each participant's attention level or capabilities, it is difficult to have clarity on the efficacy of this category. Additionally, the other possibility for the large deviation is the manner in which the study needed to be performed. Since observations needed to be completed via video chat and on varying hardware, some individuals experienced technical issues, like stuttering during the experience. This may have been a potential factor in breaking the focus of a participant. However, since the averages in the other subcategories were high with low deviations and overall the entire scale had a higher average (4.32), this subcategory may not hold as much weight at the others, and either may not be needed, or may need to be modified to eliminate potential discrepancies.

Likewise, the presence scale (*IPQ*) [23] indicated a similar trend. Although the average was above an acceptable range (3.92), like the *attentional focus*, it too suffered a high deviation of 1.25. Also, like *attentional focus*, it is likely that this large deviation was also a product of technical issues. Recall that the four statements used for this scale were:

- In the experience I had a sense of "being there"
- I felt present in the virtual space

- The virtual world seemed more realistic than the real world
- I was not aware of my real environment

These statements are based on the physical presence the participant perceives in VR, and the lack of awareness of their real environment. As mentioned previously, one of the technical issues experienced by some participants was stuttering. This was likely due to participants using varying headsets and graphics cards, as well as having to live stream the experience. Regardless of cause, this would have an impact the user's perception of presence as it breaks the sense of "being there". Additionally, as the observations were conducted in the participants' homes, they had varying physical space in which to move. As this experience was created to move around in a large space, various participants had less room in which to explore; ultimately running out of room and thus become "aware" of their real environment's limitations. To lower the deviations in these scores, these environmental variables must be eliminated.

The suspense scale (*SS*) demonstrated both a lower average (3.35) and ahigh deviation (1.23). The high deviation of the scale may indicate that the scale may need to be modified further, or that there is a discrepancy in the actual wording of the scale. It is possible that some individuals may perceive the concept of suspense differently than others. Additionally, the use of the words "worry" and "anxious" may be a cause for confusion. Therefore, further research needs to be completed to assess the efficacy of this scale. However, the lower average of the scale indicates a problem with the project itself. The story and the project did not contain clear moments of suspense, and therefore it may have been difficult to identify them. As suspense is an important factor and is closely linked with the concept of curiosity [30], clearer moments of suspense need to be implemented in the project and storyline.

The flow scale (FFS-2) revealed an average of 3.64 and a deviation of 1.07. To review, the concept of flow [19] is the ease in which a user arrives at a pleasant optimal performance. Flow comprises eight specific factors: challenge activity; merging of acting and awareness; clear goals; direct immediate feedback; concentration; a sense of control; loss of subconsciousness; and altered sense of time.

From this perspective, the scale coincides with the eight factors quite well. However, the larger deviation and lower average indicate problems. In past studies, the flow scale was generally used for game-based interactive digital applications [31, 32] as the factors for flow were easier to implement and measure. As this project was a cinematic experience with a linear storyline, there were sufficient opportunities to create flow based on all of these factors. There was no direct instruction or clear task given during the experience for the user to be focused on, nor was there any gauge on which the user could evaluate their own performance. While participants were able to accurately report such statements as "I acted spontaneously and automatically without having to think", the task orientated statements had the largest variation in answers. This is because there was no clear task, and there was no clear task because the participants' actions were unable to affect the storyline in a linear story. This would indi-

cate that either the flow scale is ill suited to cinematic experiences with linear stories, that it needs to be heavily modified to fit this genre, or that the project needs to find a better way to apply this scale without sacrificing its structure.

The curiosity scale (CS) performed well, with an average of 4.63 the low deviation of .60. However, it is important to note that the scale only had three statements that were simplified from the original 10. So, while it is an indication that the project did create curiosity and was able to accurately measure it, it might be pertinent to use the full-scale to get a more accurate view of the curiosity factors in the experience. Additionally, since curious types [7] were used in the project, expanding this scale to target those specific types may also prove beneficial to fine tune the results.

The enjoyment scale (ES) had the highest average of 4.75 and the lowest deviation of .43. Although it only consists of two statements, they were relatively simple, and the participants were able to answer them clearly and accurately. While it would be prudent to continue research to expand the scale, it was effective in relation to this project. This is because, in conjunction with the scale, enjoyment was also able to be observed during the experience. This enjoyment was observed objectively by the researcher in the form of participants smiling, chuckling, laughing, and some dancing.

The aesthetic pleasantness scale (APS) had an average of 4.43 with a deviation of .80. As Section 2.5 stated, this scale encompassed the elements of story-world, characters, and emotion. The high average and low deviation indicate that aesthetic pleasantness overall may play a more important role in narrative engagement than initially thought. To explain this assumption, this scale can be directly compared to the qualitative findings discussed in the next section.

4.2 Observational Analysis

First, consider the following codes: Aesthetic focus (life), aesthetic focus (environment), and focus on character. The aesthetic focus (life) was coded as such to include organic elements within the experience. These included objects such as wildlife, trees, grass, etc. Aesthetic focus (environment) included items such as the sky, the waves, and the weather. Combined, these elements make up the story-world, and part of the emotional element as stated in the APS. As mentioned in the previous section both codes had high frequencies throughout the experience, 56 and 45 respectively, with a combined total of 101. Additionally, focus on the character (which also makes up the character element of the APS) also had high frequencies with a total of 120. This implies that focusing on the aesthetics of the story-world is nearly as important as focusing on the characters in story. As all of these codes are also part of the APS, this further indicates the importance of aesthetic pleasantness overall. Consider that the frequency of these codes combined equals to 221, whereas the rest of the codes combine equals to 99, with an overall total of 320. Based on the number of frequencies for this study, participants spent nearly 70 percent of their time focusing on the aesthetics (character and story-world).

Another indication of its significance is that of verbal feedback received after completion of the story. Upon completion of the experience, each participant was asked which scenes they had an emotional connection to. With the exception of one outlier, all the other participants named the same two scenes having affected them the most. The first scene identified (Figure 6) involved the participant standing on the edge of a lighthouse at night with Northern lights in the sky, and the lights reflecting on the ocean water.

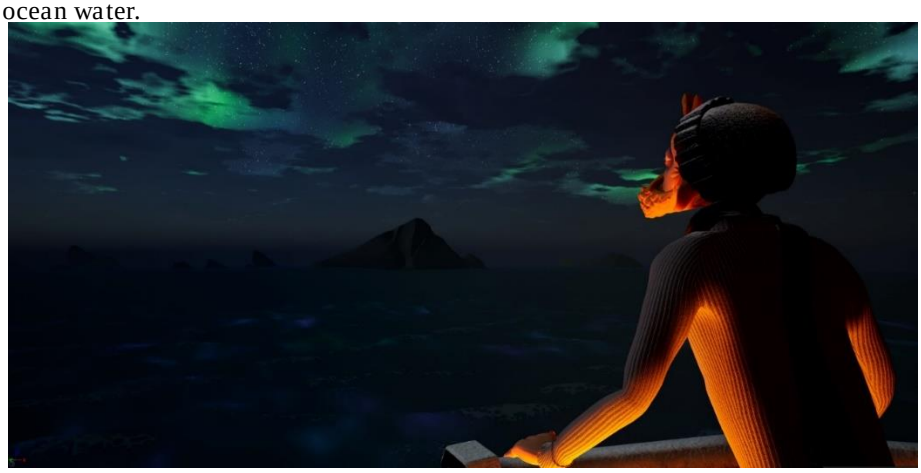

Fig.6. First Identified scene by participants

The second scene identified, involved the user being immersed in the ocean, physically flowing through the ocean, and elements and ocean creatures becoming bioluminescent (Figure 7). Both of these scenes had very strong visual attributes attached to them, which would lead to the possibility of investigating their visual attributes further.

Fig.7. Second scene identified by participants

Additionally, the one participant that chose a different scene, chose a scene involving the characters gardening together (Figure 8). The participant stated that this scene made them feel "nostalgic". This coincides with the postulation that aesthetic content could relate to the personal background and previous experiences of the recipient, evoking congruent feelings in the participant [28
].

Fig. 8. Outlier scene identified by participant

The next data to analysis is the interact with shell (meaningful) code. This code was attached to the specific interaction with an object (shell) during the experience. This object was the only object directly referenced in the experience's narration. Furthermore, it was also used as a physical representation for a "moment of unexpected change" [3] and as an ignition point for the story [3]. It was for these reasons that the shell was deemed a meaningful item [33]as meaning plays a dominant role in guiding attention in scenes of stories. As mentioned earlier, 100 percent of participants triggered this code. It is interesting to note that although some participants interacted with the shell more than once, all participants interacted with the shell at the same moment the story. This moment happened at the very end of the story when the character interacts with the participant. Conversely, the *interact with other objects code* had a more varied response. When comparing the two, while interact with shell (meaningful) had a higher deviation of .87, interact with other objects had a large variation of 3.37. This variation may be attributed to a few things. Firstly, this may connect directly to the sense of flow, as there was no clear direction, instruction, or task given to the participant at any time. Thus, the participant may have been unsure about what they could or should interact with. The exception of this, of course, being the shell, as this was directly used in the story. Personalities differences also may have influenced this variation, as some individuals may be more inclined to be tactile and want to touch and explore things while others may be of a more timid nature. Although these items

were used as curiosity types [7] and behavioural residue [6], it is unclear if they had any true bearing on the narrative engagement of the story as a whole and may need to be assessed individually. However, it can be postulated that the interaction with the shell had a more consistent response because it was a part of the story thus giving it more meaning, whereas the other objects were not. This may imply the that interactive objects require more meaning or purpose to the story in order to have consistent interaction and engagement.

Like the *Interact with shell* code, the *interact with character* code was largely initiated at the very end of the story at the same time for all participants. To clarify, in the final scene the character turns to the participant and gestures form them to come to them and sit down, where the interactable shell is also located. The significance of this is that although the participant had multiple opportunities to interact with the character, 90 percent of them only did so at the end when the character interacted with them first. This may signify that for a user to engage with an NPC, the NPC must first engage with them.

Focus on crow and *follow character's gaze* were both diegetic devices within the experience to gain the attention of the participant and engage them. The crow was purposely made as a focusing diegetic device, while the gaze of the character was an accidental addition. The crow was first introduced in the first scene where the participant could interact with it, and it would appear throughout other scenes using a sound cue to direct the focus of the participant. This proved to be a semi-accurate way to direct focus, as all participants were able to focus on the crow at one point or another. However, the results were not very consistent, which may be because the crow is not a part of the story and holds no other significance.

The *following character's gaze* code was accidental, as it was a product of the character's natural personality. To clarify, the code was initiated whenever the character would point while looking at something, looked out to sea, or was otherwise searching for something. The participant would then follow the gaze and direction of the character. This signifies engagement and connection to the character as well as curiosity, as the user is trying to physically look where the NPC is looking. While the frequencies of this are varied, it appears to add to the engagement of the story, as it is a more natural occurrence than a random appearance of a bird. If purposely controlled, it may prove a more effective device in gaining a keeping attention on the story.

5 Conclusion

5.1 Summary of Research Data

To review the quantitative data, the narrative engagement scale worked moderately well, however the subcategory of attentional focus must either be eliminated or modi-

fied to better adapt to cinematic VR experiences and eliminate possible discrepancies due to different cognitive abilities. The flow scale may also not be well suited to certain VR cinematic experiences that follow a linear storyline, as users do not have a specific task assigned to them or have the ability to influence the outcome of the story. The alternative to this is that the gameplay itself would have to change in order to adapt to the concept of flow. The curiosity scale worked successfully and with a reasonable degree of accuracy but was relatively simplistic. Further evaluation and research are needed to develop a more in-depth scale regarding narrative engagement. This also applies to the scale of suspense, with the addition of requiring further development and research on the relationship between curiosity and suspense, as well as to the practical implementation of opportunities to create suspense within the experience.

Regarding qualitative data, aesthetic pleasantness appears to play a significant role in narrative engagement, and therefore needs to be expanded and further explored in depth. Additionally, interactive items may need to hold more meaning for them to be interacted with consistently. Using a diegetic item to focus the attention of the user is potentially an effective way to assist with engagement, but the focus needs to be more purposeful. Finally, meaningful interactions with NPC characters may be dependent on the NPC character that initiates interaction first, but the user is more likely to engage in mirroring the behaviour of the NPC i.e., looking where they are looking.

5.2 Limitations

As this study was conducted via internet, participants needed to have access to their own headsets and VR compatible PCs. This greatly reduced the potential number of participants to only those who had a specific brand of headset. Additionally, although there was nearly double the number of people who expressed interest in the study, only half followed through. This is likely due to the observational requirement needed over Zoom, which some participants were unwilling to do. A small sample size is problematic as in increases the bias and lacks the statistical power to find significant effects in an overall population.

5.3 Recommendations

A larger sample size would give a more accurate representation of populous and eliminate many deviations in quantitative and qualitative data while increasing its validity. Additionally, the study needs to be performed in a more controlled environment. This means the environment needs to use the same hardware, headset, the same graphics card, and have the same room scale. It is also inadvisable to conduct such an experiment online as it introduces other technical issues, such as stuttering or prolonged delays, as well as the inability to accurately see the entirety of the participant's body during observation.

Furthermore, the VR project needs to have more opportunities for suspense, and more research should be explored on other suspense scales, creating suspense, and its definition. This was lacking in both the research and project, and as it is linked with curiosity [30] and as curiosity largely impacts narrative engagement [8, 4, 34], it would be beneficial to have a more in depth understanding of it. Along with suspense, the curiosity scale would also benefit from more exploration into its assessment, concepts, and the relationship of curiosity to narrative engagement as a whole. This would provide a well-rounded data set, increasing accuracy and validity. Finally, additional research should be conducted on the importance of aesthetics in cinematic VR experiences, and aesthetic scale needs to be modified and expanded based upon those recommendations.

The use of the flow scale is probably not appropriate for cinematic VR experience with the linear storyline therefore either needs to be eliminated from the narrative engagement measurement or heavily modified to better fit with the genre. Likewise, the attentional focus aspect of the narrative engagement scale also either needs to be eliminated or heavily modified to eliminate discrepancies based upon potential cognitive differences, capabilities or personalities.

References

1. Wolfe, A. Louchart, S. and Loranger, B. (2022). The Impacts of Design Elements in Interactive Storytelling in VR on Emotion, Mood, and Self-reflection. 10.1007/978-3-031-22298-6_40.
2. Koenitz, H. *et al.* (2018) 'The Myth of `Universal' Narrative Models', in Rouse, R., Koenitz, H., and Haahr, M. (eds) *Interactive Storytelling*. Cham: Springer International Publishing, pp. 107–125
3. Storr, W. (2019) *The Science of Storytelling*. London, UK: William Collins.
4. Richardson, D. C. *et al.* (2018) 'Measuring narrative engagement: The heart tells the story', *bioRxiv*. doi: 10.1101/351148
5. McCrae, R. R., Gaines, J. F. and Wellington, M. A. (2012) 'The Five-Factor Model in Fact and Fiction', in *Handbook of Psychology, Second Edition*. doi: 10.1002/9781118133880.hop205004
6. Gosling, S. (2008) *Snoop*. London, UK: Bookmarque.
7. To, A. *et al.* (2016) 'Integrating Curiosity and Uncertainty in Game Design', *Proceedings of 1st International Joint Conference of DiGRA and FDG*, pp. 1–16
8. FEARGHAIL, C. O. *et al.* (2018) 'Director's Cut - Analysis of Aspects of Interactive Storytelling for VR Films', in Rouse, R., Koenitz, H., and Haahr, M. (eds) *Interactive Storytelling*. Cham: Springer International Publishing, pp. 308–322
9. Brown, S. A. *et al.* (2020) 'Capturing User Emotions in Interactive Stories: Comparing a Diegetic and a Non-diegetic Approach to Self-reporting Emotion', in Bosser, A.-G., Millard, D. E., and Hargood, C. (eds) *Interactive Storytelling*. Cham: Springer International Publishing, pp. 229–242
10. Bhide, S., Goins, E. and Geigel, J. (2019) 'Experimental Analysis of Spatial Sound for Storytelling in Virtual Reality', in Cardona-Rivera, R. E., Sullivan, A., and Young, R. M. (eds) *Interactive Storytelling*. Cham: Springer International Publishing, pp. 3–7
11. Busselle, R. and Bilandzic, H. (2008) 'Fictionality and perceived realism in experiencing stories: A model of narrative comprehension and engagement', *Communication Theory*, 18(2). doi: 10.1111/j.1468-2885.2008.00322.x
12. Bindman, S. W. *et al.* (2018) 'Am I a bunny? The impact of high and low immersion platforms and viewers' perceptions of role on presence, narrative engagement, and empathy during an animated 360° video', in *Conference on Human Factors in Computing Systems - Proceedings*. doi: 10.1145/3173574.3174031
13. Schoenau-Fog, H. (2011) 'Hooked! -- Evaluating Engagement as Continuation Desire in Interactive Narratives', in Si, M. *et al.* (eds) *Interactive Storytelling*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 219–230
14. Knobloch, S. *et al.* (2004) 'Affective news: Effects of discourse structure in narratives on suspense, curiosity, and enjoyment while reading news and novels', *Communication Research*, 31(3). doi: 10.1177/0093650203261517
15. Hartmann, T., Stuke, D. and Daschmann, G. (2008) 'Positive Parasocial Relationships with Drivers Affect Suspense in Racing Sport Spectators', *Journal of Media Psychology*. Taylor & Francis Group, 20(1). doi: 10.1027/1864-1105.20.1.24
16. Roth, C. (2015) *Experiencing Interactive Storytelling*.
17. Spielberger, C. D. *et al.* (1979) 'Preliminary manual for the state-trait personality inventory (STPI)', *Unpublished manuscript University of South Florida Tampa*

18. Spielberger, C. D., Peters, R. A. and Frain, F. S. (1981) 'Curiosity and anxiety', in *Curiosity research: Basic concepts and results*

19. Csikszentmihalhi, M. (1997) 'Finding flow: the psychology of engagement with everyday life', *Choice Reviews Online*, 35(03). doi: 10.5860/choice.35-1828

20. Jackson, S. A. and Eklund, R. C. (2002) 'Assessing flow in physical activity: The flow state scale-2 and dispositional flow scale-2', *Journal of Sport and Exercise Psychology*, 24(2). doi: 10.1123/jsep.24.2.133

21. Jackson, S. A., Martin, A. J. and Eklund, R. C. (2008) 'Long and short measures of flow: The construct validity of the FSS-2, DFS-2, and new brief counterparts', *Journal of Sport and Exercise Psychology*, 30(5). doi: 10.1123/jsep.30.5.561

22. Slater, M., Usoh, M. and Steed, A. (1994) 'Depth of Presence in Virtual Environments', *Presence: Teleoperators and Virtual Environments*, 3(2). doi: 10.1162/pres.1994.3.2.130.

23. Schubert, T., Friedmann, F. and Regenbrecht, H. (2001) 'The experience of presence: Factor analytic insights', *Presence: Teleoperators and Virtual Environments*, 10(3). doi: 10.1162/105474601300343603.

24. Witmer, B. G., Jerome, C. J. and Singer, M. J. (2005) 'The factor structure of the Presence Questionnaire', *Presence: Teleoperators and Virtual Environments*. doi: 10.1162/105474605323384654.

25. Schwind, V. *et al.* (2019) 'Using presence questionnaires in virtual reality', in *Conference on Human Factors in Computing Systems - Proceedings*. doi: 10.1145/3290605.3300590.

26. Vorderer, P., Klimmt, C. and Ritterfeld, U. (2004) 'Enjoyment: At the heart of media entertainment', *Communication Theory*, 14(4). doi: 10.1111/j.1468-2885.2004.tb00321.x

27. Sweetser, P. and Wyeth, P. (2005) 'GameFlow: a model for evaluating player enjoyment in games', *Computers in Entertainment (CIE)*, 3(3). doi: 10.1145/1077246.1077253

28. Cupchik

29. Rowold, J. (2008) 'Instrument Development for Esthetic Perception Assessment', *Journal of Media Psychology*. Taylor & Francis Group, 20(1). doi: 10.1027/1864-1105.20.1.35

30. Hoeken, H. and Sinkeldam, J. (2014) 'The Role of Identification and Perception of Just Outcome in Evoking Emotions in Narrative Persuasion', *Journal of Communication*. Blackwell Publishing Ltd, 64(5), pp. 935–955. doi: 10.1111/jcom.12114

31. Isbister, K. (2006) *Better game characters by design: a psychological approach*, *Education*

32. Isbister, K. (2018) *How Games Move Us*, *How Games Move Us*. doi: 10.7551/mitpress/9267.001.0001

33. Henderson, J. M. and Hayes, T. R. (2017) 'Meaning-based guidance of attention in scenes as revealed by meaning maps', *Nature Human Behaviour*, 1(10). doi: 10.1038/s41562-017-0208-0

34. Loewenstein, G. (1994) 'The Psychology of Curiosity: A Review and Reinterpretation', *Psychological Bulletin*, 116(1). doi: 10.1037//0033-2909.116.1.75