Utrecht University

Adaptive Textual Guidance using Specificity in Serious Virtual Reality Games

Ioannis Fermanis
ICA-6511325

January 20, 2021

MSc-program in Game and Media Technology (GMT)
Department of Information and Computing Sciences
Graduate School of Natural Sciences (GSNS)
Utrecht University

INFOMGMT1 (15 ECTS)
INFOMGMT2 (25 ECTS)

Supervisors:

Dr. Egon L. van den Broek
Dr. Herre van Oostendorp
MSc. Chris Ferguson

Adaptive Textual Guidance using Specificity in Serious Virtual Reality Games

Ioannis Fermanis
6511325

January 20 2021

Abstract

Virtual Reality (VR) games have become increasingly more popular in the digital age. Educational games can benefit from this by making use of VR technologies and bestow an even better experience than previously. Having said that, certain elements can reduce the overall VR experience, such as being lost in a task finding game. This research assesses how adaptive textual guidance can be utilized for the purpose of reducing cognitive overload and enhancing engagement, presence, interest and knowledge retention in virtual environments by adapting to the player's needs. The lostness measure system developed in this paper is applied to the PlayStation 4 Virtual Reality (PS4VR) game "The Chantry" where the difficulty of finding items is adjusted in real-time, based on how the player is performing. Experiments ware conducted comparing 2 groups where one played the normal version of the game, while the other played a version that includes the adaptive measure. 31 participants took place in the experiments, split into two groups of 17 and 14. Each participant filled out a questionnaire and a knowledge test after their play-through of the game, measuring their interest, presence, engagement, cognitive load, story knowledge, and spatial knowledge. Results show that the adaptive system worked as intended, reducing and increasing specificity only when deemed necessary. Results revealed that participants playing with the adaptive version scored higher on both story and spatial knowledge, while having a reduced cognitive load. No differences were found in presence, interest, and engagement. Considering all this, ensuring a more overall intermediate experience for exploratory VR games will enhance the players knowledge retention, spatial information retention and reduce their cognitive load.

1 Introduction

Games have proved to be a powerful tool in teaching and learning [1, 2, 3, 4, 5, 6, 7]. One of the key benefits of games and virtual reality (VR) is that they can be immersive, thus making the player further motivated to learn more about the world they are experiencing. Meaning that any method that boosts the player's experience would be a potent benefit for any game. An interesting way to approach this would be to adjust the difficulty of the game based on the player, considering it has been proven that experience will drop if the player is having a hard or too easy of a time in the game [1, 8].

Admittedly, for more complex games it is potentially quite challenging to ensure that all players will play the game on the difficulty that is most suited specifically for them. This signifies that a method that adapts in real-time while the player is still playing would be an extremely convenient solution. As certain types of games provide instructions in order to describe how to complete an objective, making these more specific will make these tasks easier and vice versa. The proposed adaptive algorithm is implemented based on theories from cognitive psychology, information retrieval, and the theory of specificity. Experiments were carried out in this paper to examine the algorithms capabilities by measuring the player's cognitive load [9], engagement, interest, presence, knowledge and spatial information retention of the game. This is done by having them fill out a user questionnaire/survey [6, 10, 11].

There are a few terms that require explanation before proceeding. "Performance" and "Efficiency" are defined as how fast can the player locate a target, as well as the ability to retain information. Additionally, being "lost" is defined as giving up before finding the target even if the user is on the right track. Lostness refers to how lost a user is when trying to locate a target. But this is not a subjective user feeling since lostness is a degradation of their performance [12]. Lostness can be calculated as long as the number of information items the player has inspected and the number of required items that are required to be inspected is known. "Interest" refers to how interested the player is in the game and "engagement" how psychologically engaged they are while playing. "Experience", in this case, is a general term containing presence, interest, engagement and immersion all at the same time. The last significant definition that needs mentioning is the relation between "immersion" and "presence" in virtual environments (VE). Mestre [13] says "immersion is intended to instill a sense of belief that one has left the real world and is now "present" in the virtual environment. This notion of "being present" in the virtual world has been considered central to VE. Thus, whereas immersion is a "technology-related", objective aspect of VEs, presence is a psychological, perceptual, and cognitive consequence of immersion. Presence is thought of as the psychological perception of "being in" or "existing in" the VE in which one is immersed" [13]. This study focuses on the psychological perception aspect by manipulating difficulty and not the technological one, therefore immersion is not measured.

Cognitive load is related to the amount of information being presented to the user [9]. It has been proven that reducing the player's cognitive load will maximize the user's performance and satisfaction, as well as their ability to retain information [1, 14]. This occurs when users do not have a well-defined goal that matches the information presented, causing them to spend "useless" time trying to get acquainted with this newfound information. However, reducing cognitive load by too large of a degree can lead to cognitive underload, which is detrimental to the player's learning experience [15]. Another major factor affecting performance is how lost the user is. It has been shown that lostness is highly related to time and efficiency [1]. This is also in line with research revealing that reducing lostness or providing guidance when a user is lost is beneficial [16]. When assisting the player, performance rises, but at the same time, if a player feels like they require assistance then the "sense of presence" is broken [1, 17]. This suggests that it would be best to preemptively help the player before they start feeling like they need help or even quitting.

Several ways can be used to lower the difficulty and increase efficiency when searching for a target among distracters. Increasing the N-N (Non-target and Non-target similarity), lowering the T-N (Target and Non-target similarity) will result in the best efficiency and vice versa for worst [18, 19] Furthermore, increasing T-N does not matter when N-N is already high and decreasing N-N does not matter when T-N is already low. Also, when N-N is low then T-N is even more potent. Furthermore, searching for an item "A" among a sea of "B" can be harder than searching for item "B" in a sea of "A" [20]. It has also been shown that color is effective at making a target stand out [21], even more so than shape [22]. The best colors to utilize are red, green and yellow [20]. Luminosity is another excellent way of making a target stand out [23, 24] but opacity is not [20]. Set size also affects performance [23] but the range of distracters does not matter [18] and search efficiency is not correlated to serial or parallel search [19]. It has also been proven that showing an image of a target is more effective than describing the target through text form since the amount of helpful information that can be provided, in the latter case, is limited [17, 24, 25]. Finally, the human mind categorizes certain shapes and objects and searches a target with those categories in mind, instead of using the actual shape. An example would be where a 10° tilted line is interpreted the same as a 15° tilted line angle since they are categorized as the same in the human mind [18, 20, 22]. One last finding in terms of search is that when users are asked to confirm whether the target exists or not they are more efficient when the target does indeed exist [20, 26].

It is important to note that the "intention" of the guide is significant and will play a factor in influencing the search of the user [27]. It has been found that search instructions and guidance can override the salience of objects from the player's perspective [24]. More specifically, the type of information given can influence the player's eye path and fixation, therefore deciding beforehand which order the player will perceive certain elements and characteristics of an object or environment [9]. The specificity of the instructions can also influence which characteristics will the player find more important when observing an object or environment [28], effectively deciding which parts will be observed and which ones will not. If both general and specific information is given, the player will place a larger emphasis on the specific one and will likely ignore details of the general one. Additionally, this will inevitably control how and what type of information the player will recall afterward.

In this study, specificity labels are used, where they guide the player towards the next in-game objective or location of interest. In the adaptive version, the labels will be changing difficulty after every objective, becoming more specific to lower the difficulty or more vague to increase the difficulty, based on the player's performance in the previous objective. But the first label cannot possibly be adaptive. The reason being, implementing an adaptive system without prior knowledge of the nature of the interaction is impossible [29]. Another relevant discovery in regards to the knowledge test is that even if people can locate the target easily this does not necessarily mean that they will remember details about it or the environment [26].

There is evidence supporting the claim that educational games are most effective when they are immersive and engaging, as well as when the player is motivated towards finishing the game [30]. There is also evidence showing that specifically recall and long-term knowledge retention increases when immersion and engagement are high [6]. More specifically for VR, research reveals that VR games improve the player's spatial learning even when compared to immersive desktop games [7]. On the other hand, it has been shown that player experience can be hindered due to them being lost in the game, not knowing where their next objective or target should be [22]. This, combined with the evidence that adaptive games have potential, supports the claim that creating an adaptive game specifically for guiding a player should result in the better overall performance of the game's ability to instill information to the player, with the possibility of enhancing the player experience as well.

2 Lostness Formula & Specificity Labels

This paper is following Pauline A. Smith's [12] formula and definition of lostness where it is viewed as a term related to the degradation of player performance instead of a purely subjective feeling of being lost.

2.1 Lostness formula

The player moves in-game by traversing through nodes. These nodes are specified points where the player is allowed to move towards discretely. By logging the player's movement through these nodes, the algorithm can output the player's lostness. This is done by calculating the number of nodes between the player and the target, and recording whether each move brings the player closer or farther to the target.

More specifically, the implemented lostness algorithm is based on Equation 1.

$$[h]\sqrt{\left(\frac{N}{S}-1\right)^2+\left(\frac{R}{N}-1\right)^2}, \tag{1}$$

where R is the minimum number of nodes needed to be visited and reach the target, S is the total number of nodes a player has visited, and N is the number of unique nodes a player has visited. Unique nodes are defined as nodes the player has yet to visit. This formula will return a value between 0 and $\sqrt{2}$. 0

indicates that the task has been completed by traversing the least amount of nodes possible and the player was not disoriented (high navigational efficiency). $\sqrt{2}$ indicates that the player was completely disoriented (low navigational efficiency) and has unnecessarily traversed through multiple nodes. The S and N-values will be recorded as the player carries out the task whilst the R-value is calculated in real time every time a new task is completed.

But instead of using the normal lostness formula, the local lostness formula will be used. Reason being, for this experiment, each task has several objectives that can be completed in any order, so rather than only focusing on these sets of objectives that belong to each task, local lostness takes into consideration all the objectives, essentially treating the whole game as one long task [14].

Equation 2 for local lostness can be derived by using the Equation 1. Here, R, S, and N-values of each objective (O) being summed together to give a measure of lostness for the full game:

$$[h]\sqrt{\left(\frac{\sum_{o=1}^{n} N_o}{\sum_{o=1}^{n} S_o} - 1\right)^2 + \left(\frac{\sum_{o=1}^{n} R_o}{\sum_{o=1}^{n} N_o} - 1\right)^2}, \tag{2}$$

with n representing the overall number of objectives O.

In this measure, the S and N-values are reset to 0 after each completed task. Once again, the R-value has to be calculated for every new task whilst S and N-values are constantly being recorder throughout the entire play-through.

The R, S, and N-values from each objective are fed into the original lostness formula to give a full-game local lostness measure LL. This reflects the full path taken by a player throughout the whole game. Finally, it is important to note that the intention of the system is to provide an intermediate experience, meaning that most players will be falling close to the middle gray area of lostness (0.5>LL>0.4) where this suggests that they are both "about to be lost" and "about to stop being lost" at the same time [12].

2.2 Specificity Labels

The game in question "The Chantry" describes the story of Dr. Edward Jenner and how he invented the vaccination against the smallpox virus. In order to progress through the game, the player needs to explore the house of Jenner. The payer will find locked doors and window shutters, blocking them from progressing. when attempting to unlock them, a list of objectives will appear. These objectives represent a particular topic, or episode, of the story (see **Figure 2**). Each list is a task where players must find items that contain story information, in the form of an audio narrative, which must be found in order to unlock the aforementioned doors and window shutters. To do this, players must pick up items and rotate them around, searching for text written on these items. The text on a mandatory item is always the same as the name of the current chapter (see **Figure 1**).

As explained earlier, specificity labels will be used to guide the players throughout "The Chantry". The labels will have text on them explaining where

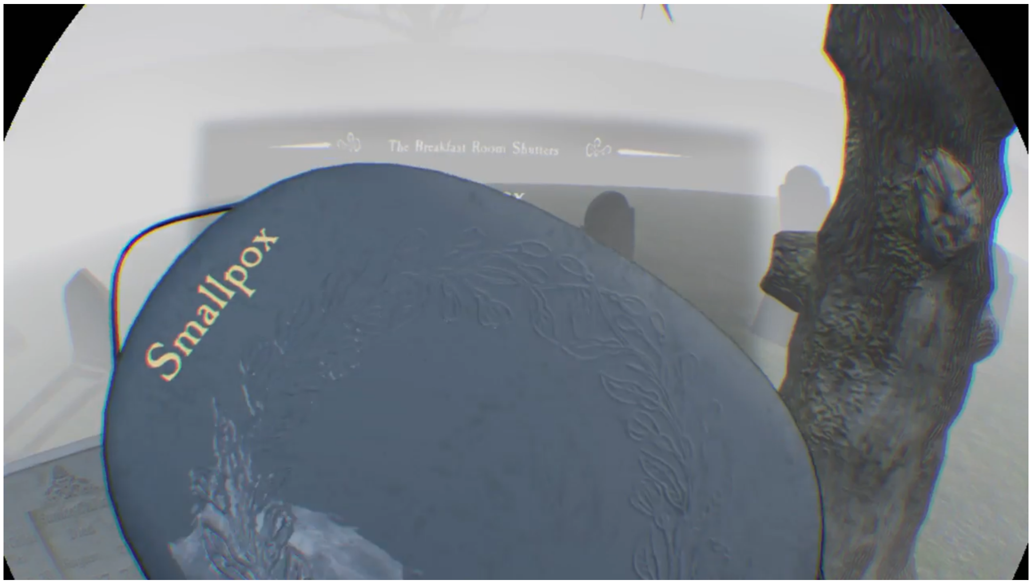

Figure 1: Player discovering one of the items necessary to progress the "Smallpox" story chapter

the next location of interest or item should be. The explanation on the labels will become more vague as the game becomes harder, or more specific as it becomes easier, depending on the player's needs. There are 5 difficulty stages ("Very Easy"-"Easy"-"Normal"-"Hard"-"Very Hard"). The difficulty changes one level down or up based on the player's performance from the previous objective. The format for every difficulty level is as follows. "Very Easy" contains specific information related to the location of the objective and the item itself that needs to be picked up (target item). "Easy" contains vague information related to the location of the objective but still contains specific information about the target item. "Normal" only contains specific information regarding the target item. "Hard" informs about the target item vaguely. "Very Hard" only describes the target item in a very generic way. The difficulty of the non-adaptive version does not change and is constantly on "Normal" (see **Figure 2**).

3 Methods

The purpose of this study is to test whether providing adaptive textual guidance will enhance the player's engagement, interest, presence, knowledge retention and affect cognitive load in a VR environment. This is achieved by implementing the lostness formula which determines if the player is lost and requires assistance.

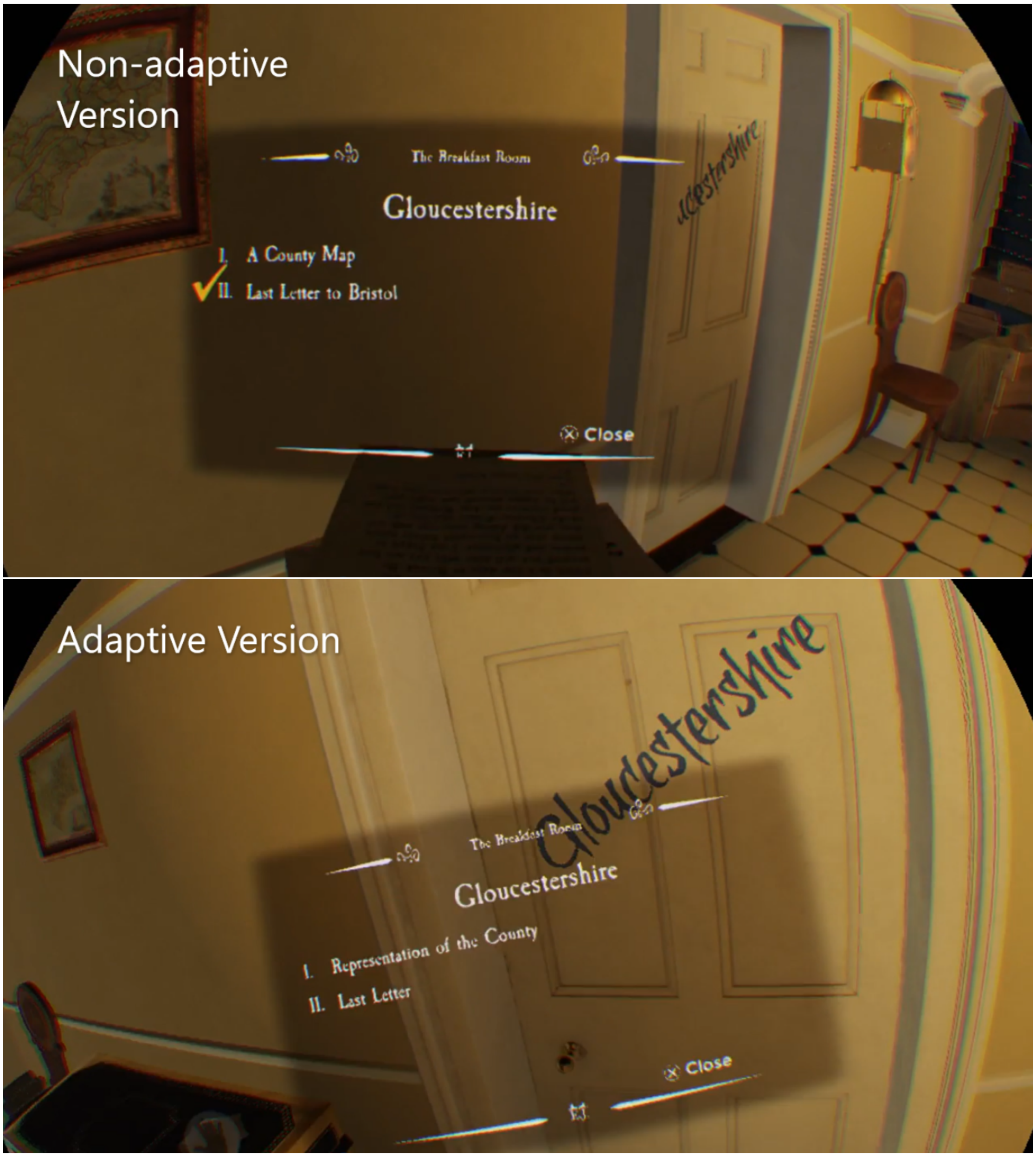

Figure 2: Adaptive and non-adaptive Versions. The labels are less specific (harder difficulty) in the adaptive version in this particular instance

3.1 Participants

31 participants were included in this study. 16 were male and 15 were female. All had normal or corrected-to-normal vision. 2 were well versed with VR, 11 had some previous experience and 17 did not have any previous experience (Average=1.8; 1=No experience; 5=Well versed). All participants were students from Utrecht University and all had at least an intermediate level of English. Participants with Epilepsy, severe migraine, or nausea were not allowed to participate or continue the experiment and therefore were not included in this study. Fortunately, no such case occurred.

3.2 Materials

The VR game: "The Chantry" was developed by Steel Minion Studios and was played on a Playstation 4 console using a Sony Playstation 4 Virtual Reality headset and a PlayStation 4 DualShock wireless hand controller. The game was developed on the PhyreEngine and has been modified with the adaptive system for the purpose of this experiment. The programming language used is C++ through the use of Microsoft Visual Studio 2017. A knowledge test and a questionnaire were given post-test to gauge the participant's story knowledge and spatial information retention of the contents of the game, alongside their engagement,interest and presence.

3.3 Experiment Variables

The independent variable of this experiment is the condition (either adaptive or non-adaptive version of the game). The dependent variables are the knowledge test answers (spatial and story questions), the experience questionnaire answers (presence, interest and engagement) and cognitive load.

3.4 Design

This experiment uses a between-groups design. Two groups are formed, group A plays through the game without the adaptive algorithm, while group B plays the game with it. The number of in-game tasks for both groups is the same, as well as the knowledge test and questionnaire. This design procedure was chosen in order to measure the impact of the algorithm [4, 6].

3.5 Procedure

As per the terms and conditions followed in regards to COVID-19 by Utrecht University, the researcher thoroughly cleans the PS4VR headset and the PlayStation 4 DualShock controller with sanitizer. Throughout the entirety of the experiment, a 1.5 meter distance will be kept between the researcher and the participant. The participant is asked to sign a written consent regarding the experiment. Following this, it is explained to the participant that this is an educational game aiming to provide an informative experience about the early days of vaccination of the smallpox virus. The participants will be answering a knowledge test and a questionnaire based on the game after they finish playing it. Additionally, they are instructed with the knowledge that the researcher will not be interfering with the participant or provide any help with clearing the game.

The researcher explains that first there will be a 10 minute training session in the game "A night in the forum" as a means to get the participant familiar with the controls, and then playing the main game "The Chantry". Subsequently, the researcher does indeed let the participant know about the 30-minute playtime limit when playing "The Chantry" and how they will be asked to stop once they reach it. Afterwards, the researcher explains how to equip the headset and will continue to provide help verbally to the participant until the headset is in a comfortable position. Next, the participant is handed the PS4 hand controller and before starting the game the researcher informs the participant that in-game dialogue cannot be skipped and it has significance in the knowledge test.

Now the participant can start playing the game. In-game, notifications will pop up, the specificity labels. These labels provide hints on what the player should be doing or be looking for, in order to complete the current task. Once the player has completed the current task, the specificity label for the next task will appear. It is possible to complete some of the tasks out of order. For the participants in group B, depending on how well or poorly they are performing, the algorithm will be adjusting the difficulty of the game and adapting to the player's lostness. As explained earlier, there are five difficulty modes (Very Easy - Easy - Normal - Hard - Very Hard). If the player performed well in the previous task, finding the right items and location efficiently without making many unnecessary steps, then the difficulty will shift one stage towards harder. This changes the label for the next task making it less informative, essentially making the task harder. The opposite is also true when the player is performing poorly. The very first objective will always start at "Normal", since it is not possible to know beforehand what the right difficulty for this particular player is [29]. For group A, the game difficulty will not be shifting on "Normal" throughout the entire game.

Once the participant is done, or once the timer is up before they complete the game within 30 minutes then the researcher will stop the participant and proceed with the knowledge test (24 questions), along with the questionnaire (43 Overall Questions; 14 Presence; 19 Engagement; 10 Cognitive Interest). Each knowledge test question can be answered by selecting true or false (0 - 1) and for the questionnaire one of the five options ranging from 1 - 5. The knowledge test will be testing the participant's retention of spatial information, as well as the in-game presented story information related to the Smallpox [6]. The questionnaire will be gauging the participant's presence, engagement and cognitive interest. The duration for the entire experiment is 60 minutes.

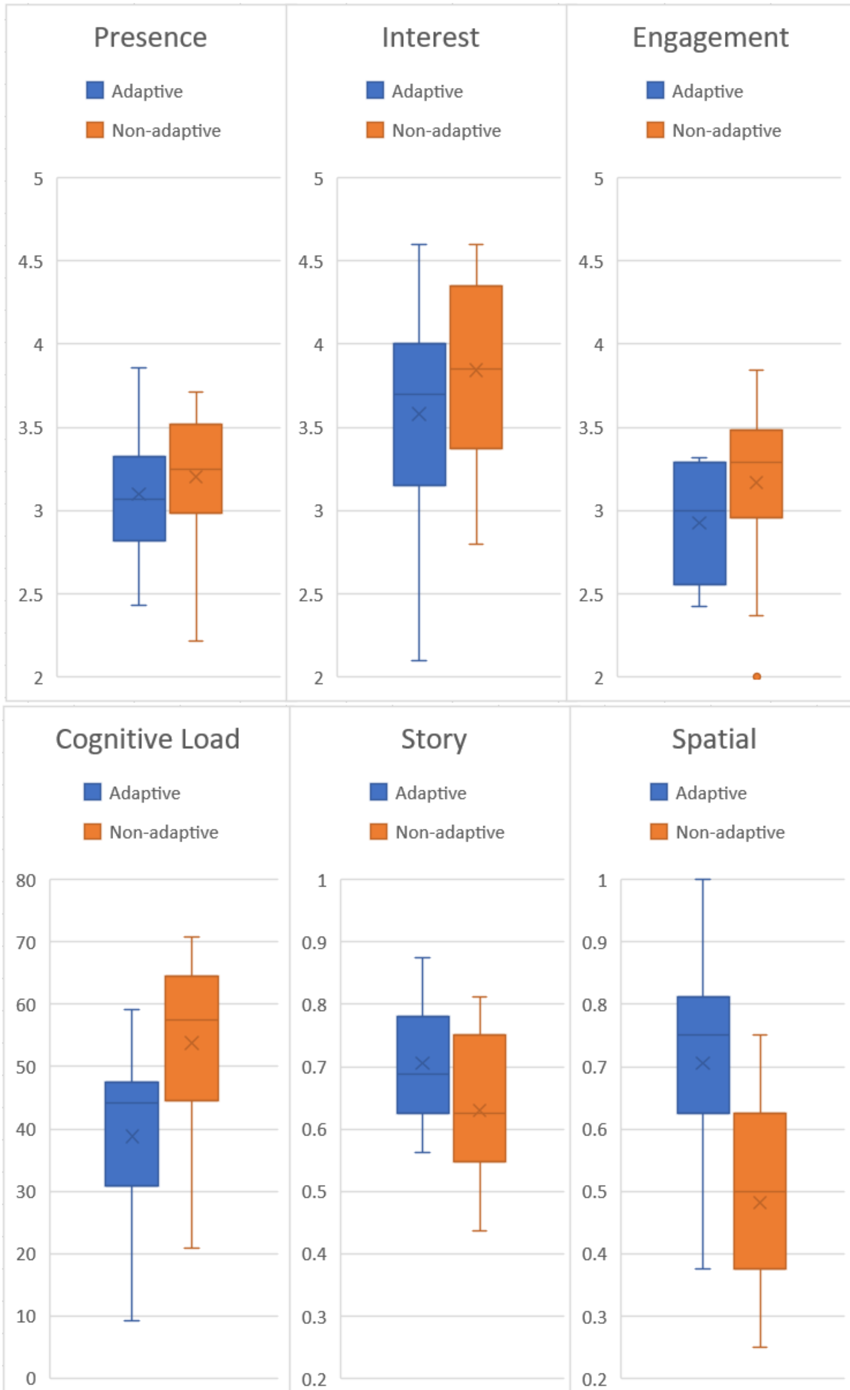

Figure 3: Presence, Interest, Engagement, Cognitive Load, Story Questions and Spatial Questions

Variable (Range)	Adaptive Mean (SD)	Non-adaptive Mean (SD)	$F(6,24)$	p	η_p^2
Presence (1-5)	3.10 (0.39)	3.20 (0.39)	0.536	0.47	0.018
Interest (1-5)	3.58 (0.71)	3.84 (0.57)	1.233	0.28	0.041
Engagement (1-5)	2.93 (0.36)	3.17 (0.49)	2.523	0.12	0.080
Load (0-100)	38.8 (13.3)	53.8 (13.7)	9.610	0.004	0.25
Story (0-1)	0.71 (0.09)	0.63 (0.11)	4.353	0.046	0.13
Spatial (0-1)	0.71 (0.17)	0.48 (0.16)	13.825	0.001	0.32

Table 1: Descriptive & MANOVA Univariate test results

4 Results

Table 1 shows a comparison between the two different conditions (Adaptive and Non-adaptive). These two groups do not share participants from one another, therefore the two groups are independent. The data is derived from 31 participants. The adaptive test contained 17 participants while non-adaptive test contained 14 participants.

Lostness is a variable manipulated by us and calculated in real time by the previously mentioned local lostness formula (see **Equation 2**), therefore we do not present it. Presence, Interest, Story questions, Spatial questions and Cognitive Load are all dependent variables which are derived from the user questionnaire. Presence refers to the sense of presence of a player in a game and Interest shows how interested the player is in the game itself, as discussed earlier. These 2 variables have a range of 1-5 (1=not at all, 5=very much so) and represent the averages of the answers in the questionnaire. Story questions and Spatial Questions are specifically the average answers for the questions that are related to story and spatial information retention respectively. These are all true/false questions and therefore have a range of 0-1 depending on how many the participant got right, with 0 being the worst score and 1 being the best. Lastly, cognitive Load refers to the amount of information a player needs to process in a singular moment, and is also measured from 0-100 (0=Cognitive underload, 100=Cognitive overload).

The box plots in **Figure 3** represent the descriptive statistics of all 31 participants. Maximum and minimum values are shown as the most top and bottom "whisker" lines respectively. The box itself shows where the middle portion of the data is located, the interquartile range. The bottom of the box represents the first quartile (Q1) and the top represents the third quartile (Q3). The middle line inside the box represents the median and "X" marks the mean. The circles are recordings that are considered outliers and therefore do not count as maximum or minimum recordings.

Following this, a normality test was conducted to prove that the above data follows a normal distribution. Considering this along with the fact the data is provided from 2 independent groups, a Multivariate Analysis of Variance (MANOVA) test was deemed appropriate in order to check whether there was a significant difference between the adaptive and non-adaptive conditions. An Analysis of Variance (ANOVA) test is statistical method of finding out if survey or experiment results are significant and determine the influence that independent variables have on a dependent variable. MANOVA is similar but can instead include multiple dependent variables which fits our case.

The following syntax is used for reviewing MANOVA results: F(hypothesis df, error df)= F; p= p-value; η_p^2= Partial Eta Squared. F is the approximate F statistic for the given effect and test statistic. Higher F values imply that the variance in the data is caused from between groups rather than within groups. Hypothesis df is the number of degrees of freedom in the model and error df is the number of degrees of freedom associated with the model errors. Then, p stands for the p-value which is used for observing whether a variable is statistically significantly different or not. $Wilk's\Lambda$ can be interpreted as the proportion of the variance in the outcomes that is not explained by an effect. Partial eta squared is the ratio of variance associated with an effect, plus that effect and its associated error variance. It shows the percentage of variance in each effect or interaction, and the error that is accounted for by that effect.

In the MANOVA, an overall significant effect in favor of adaptivity was observed ($F(6, 24) = 5,158; p = 0.002 < 0.05; Wilk's\Lambda = 0.437; \eta_p^2 = 0.563$). Afterwards, the univariate effects on the dependent variables was analyzed. More specifically Presence, Interest, Engagement, Cognitive Load, Story Questions and Spatial Questions were examined as dependent variables using a significance level of ($\alpha = 0.05$) and their results are shown in **Table 1**.

5 Discussion

It is common for players to get lost inside an exploratory game [1, 17]. Lostness is not inherently a negative attribute as it can provide a much-needed challenge when exploring, aiding in the experience of the player [3, 30]. However, the user should not be acting without a clear goal for a prolonged period of time. Players who progress aimlessly will inevitably lose interest and will therefore stop playing [8]. The problem is even more apparent in educational games as players who never manage to complete the game can never learn what the game was aiming to teach. Therefore, it makes sense that a game should not eliminate lostness entirely; but, instead, reduce it whenever necessary. By automatically adjusting difficulty in real-time through specificity labels, the game can provide help to counteract this lostness only when needed. In this paper, such an algorithm is developed and implemented in Steel Minion's PS4VR game "The Chantry". Subsequently, an experiment was carried out with two groups of participants to test the impact of this algorithm on the player. In the experiment, participants completed a play-through of the game with the adaptive system or without it, depending on their group. Following this, the participants answered a questionnaire related to their presence, interest and engagement in the game as well as questions related to the historical knowledge the game was originally aiming to provide.

A Multivariate Analysis of Variance (MANOVA) test was carried out, revealing that cognitive load and spatial and story knowledge were all positively impacted by the adaptive system. This is in line with other works showing that cognitive load does indeed influence player performance [1, 9, 11]. No difference was found in interest, engagement, or presence, implying that the adaptive system does not impact these attributes. Despite having said that, it is clear that the adaptive system increases their performance in-game as well as outside the game since participants have better knowledge and spatial information retention when the difficulty is adjusted. That leads us to believe that engagement, presence, and interest in serious games do not necessarily need to be enhanced to improve knowledge retention of the taught subject. Other experiments related to this topic seem to have arrived at the same conclusion [4].

As expected, the most sizeable difference was observed on lostness since the specificity labels combined with the adaptive system ensure that players get more/less hints when needed. This is in line with other works where it has been proven that specificity makes a difference in searching [24]. It is also in line with research showing that a lostness measure would be a valuable asset for a node-based VR game [7].

Considering all this, the answer to our original research question would be that adaptive textual guidance does not enhance the player's engagement, presence, or interest, but it does manage to enhance the player's knowledge retention, and improve the player experience by reducing cognitive load while also ensuring no cognitive underload or overload is experienced.

One of the limitations of this study lie in the methods of adjusting the difficulty. Only one factor is used, and that is specificity in the form of text labels. Another limitation is that the system only works for exploratory games, and even in that case, the formula would have to be redesigned if the game does not include a node system like "The Chantry". Having said all this, one potential future direction of this research would be utilizing other factors such as sound or haptic feedback as a means of changing the search difficulty. There has been some research done for haptic feedback previously, but no conclusive evidence was found [31]. Another future direction could be to further develop the lostness formula for a game environment without a node system, possibly measuring distance in in-game length, instead of number of nodes.

To conclude, knowledge retention in educational games is of paramount importance. Therefore, a system that enhances knowledge retention would greatly benefit any educational game. In this paper, a system is implemented in a serious VR exploratory game that improves knowledge retention by adapting the search difficulty, based on how lost the player is. While there was no increase in interest, presence or engagement, results have shown that this system aids in the teaching of players by improving knowledge and spatial information retention by ensuring a more intermediate experience for all types of players, while avoiding both cognitive overload and cognitive underload.

6 Acknowledgements

I wish to thank all the people whose assistance was a milestone in the completion of this project. I would like to pay my special regards to my supervisors Dr. Egon van den Broek and Dr. Herre van Oostendorp for providing feedback and guidance throughout the entire study, as well as Chris Ferguson, MSc. for helping and allowing me to use his programming tools along with providing the data from the experiment. I would like to express my thanks to the Steel Minions team for allowing me to use their game as the basis for this experiment. Additionally, I would like to thank my family and friends that have helped me get through these times, especially during the year 2020. Finally, I would like to thank the UU Science-Geo Ethics Review Board (SG-ERB) for approving the execution of this study.

References

[1] H. van Oostendorp and S. Karanam, "Navigating in a virtual environment with model-generated support," in *Proceedings of the 31st European Conference on Cognitive Ergonomics*, pp. 1–8, 2013.

[2] W. Doesburg, A. Heuvelink, and E. L. van den Broek, "TACOP: A cognitive agent for a naval training simulation environment," in *Proceedings of the Industry Track of the Fourth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-05)* (M. Pechoucek, D. Steiner, and S. Thompson, eds.), (Utrecht, The Netherlands), pp. 34–41, New York, NY, USA: ACM, 2005.

[3] S. Goumas, G. Terzopoulos, D. Tsompanoudi, and A. Iliopoulou, "Wordsearch, an educational game in language learning.," *Journal of Engineering Science & Technology Review*, vol. 13, no. 1, 2020.

[4] E. A. Alrehaili and H. Al Osman, "A virtual reality role-playing serious game for experiential learning," *Interactive Learning Environments*, pp. 1–14, 2019.

[5] M. Loachamín-Valencia, M.-C. Juan, M. Méndez-López, and E. Pérez-Hernández, "Using a serious game to assess spatial memory in children and adults," in *International Conference on Advances in Computer Entertainment*, pp. 809–829, Springer, 2017.

[6] L. Chittaro and F. Buttussi, "Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety," *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 4, pp. 529–538, 2015.

[7] C. Ferguson, E. L. van den Broek, H. van Oostendorp, S. de Redelijkheid, and G.-J. Giezeman, "Virtual reality aids game navigation: evidence from the hypertext lostness measure," *Cyberpsychology, Behavior, and Social Networking*, vol. 23, no. 9, pp. 635–641, 2020.

[8] C. Grappiolo, Y. Cheong, J. Togelius, R. Khaled, and G. N. Yannakakis, "Towards player adaptivity in a serious game for conflict resolution," in *2011 Third International Conference on Games and Virtual Worlds for Serious Applications*, pp. 192–198, 2011.

[9] F. van der Sluis, E. L. van den Broek, R. J. Glassey, E. M. A. G. van Dijk, and F. M. G. de Jong, "When complexity becomes interesting," *Journal of the American Society for Information Science and Technology*, vol. 65, no. 7, pp. 1478–1500, 2014.

[10] E. L. van den Broek, "On making engagement tangible," in *Proceedings of Measuring Behavior 2012: 8th International Conference on Methods and Techniques in Behavioral Research* (A. J. Spink, F. Grieco, O. E. Krips, L. W. S. Loijens, L. P. J. J. Noldus, and P. H. Zimmerman, eds.), (Utrecht, The Netherlands), pp. 90–93, August 28–31 2012.

[11] F. Meijer and E. L. van den Broek, "Representing 3D virtual objects: Interaction between visuo-spatial ability and type of exploration," *Vision Research*, vol. 50, no. 6, pp. 600–635, 2010.

[12] P. A. Smith, "Towards a practical measure of hypertext usability," *Interacting with computers*, vol. 8, no. 4, pp. 365–381, 1996.

[13] D. Mestre, P. Fuchs, A. Berthoz, and J. Vercher, "Immersion et présence," *Le traité de la réalité virtuelle. Paris: Ecole des Mines de Paris*, pp. 309–38, 2006.

[14] C. Ferguson and H. Van Oostendorp, "Lost in learning: Hypertext navigational efficiency measures are valid for predicting learning in educational virtual reality games," *Frontiers in psychology*, vol. 11, p. 3264, 2020.

[15] F. Paas, A. Renkl, and J. Sweller, "Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture," *Instructional science*, vol. 32, no. 1/2, pp. 1–8, 2004.

[16] C. Ferguson, E. L. van den Broek, and H. van Oostendorp, "On the role of interaction mode and story structure in virtual reality serious games," *Computers & Education*, vol. 143, p. 103671, 2020.

[17] K. E. Steiner and L. Voruganti, "A comparison of guidance cues in desktop virtual environments," *Virtual Reality*, vol. 7, no. 3-4, pp. 140–147, 2004.

[18] J. M. Wolfe, S. R. Friedman-Hill, M. I. Stewart, and K. M. O'Connell, "The role of categorization in visual search for orientation.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 18, no. 1, p. 34, 1992.

[19] J. Duncan and G. W. Humphreys, "Visual search and stimulus similarity.," *Psychological review*, vol. 96, no. 3, p. 433, 1989.

[20] J. M. Wolfe and W. Gray, "Guided search 4.0," *Integrated models of cognitive systems*, pp. 99–119, 2007.

[21] D. Melcher, T. V. Papathomas, and Z. Vidnyánszky, "Implicit attentional selection of bound visual features," *Neuron*, vol. 46, no. 5, pp. 723–729, 2005.

[22] J. Schmidt and G. J. Zelinsky, "Search guidance is proportional to the categorical specificity of a target cue," *The Quarterly Journal of Experimental Psychology*, vol. 62, no. 10, pp. 1904–1914, 2009.

[23] J. Palmer, "Attention in visual search: Distinguishing four causes of a set-size effect," *Current directions in psychological science*, vol. 4, no. 4, pp. 118–123, 1995.

[24] T. Foulsham and G. Underwood, "How does the purpose of inspection influence the potency of visual salience in scene perception?," *Perception*, vol. 36, no. 8, pp. 1123–1138, 2007.

[25] K. S. Jones, "A statistical interpretation of term specificity and its application in retrieval," *Journal of documentation*, 1972.

[26] H. Intraub, "Rapid conceptual identification of sequentially presented pictures.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 7, no. 3, p. 604, 1981.

[27] S. Jörg and H. Hörmann, "The influence of general and specific verbal labels on the recognition of labeled and unlabeled parts of pictures," *Journal of Memory and Language*, vol. 17, no. 4, p. 445, 1978.

[28] E. Tulving and D. M. Thomson, "Encoding specificity and retrieval processes in episodic memory.," *Psychological review*, vol. 80, no. 5, p. 352, 1973.

[29] F. Berger, A. Liapis, and G. Yannakakis, "Prototyping an adaptive educational game for conflict resolution," in *ITS 2012 Workshop: Emotion in Games for learning*, 2012.

[30] L. Wang, J.-w. Guo, C.-l. Yang, H.-s. Zhao, and X.-x. Meng, "O3d-based game learning environments for cultural heritage online education," in *International Conference on Technologies for E-Learning and Digital Entertainment*, pp. 417–428, Springer, 2010.

[31] D. Gouveia, D. Lopes, and C. V. de Carvalho, "Serious gaming for experiential learning," in *2011 Frontiers in Education Conference (FIE)*, pp. T2G–1–T2G–6, 2011.