

Improved Memory Elicitation in Virtual Reality: New Experimental Results and Insights

Joel Harman^(✉), Ross Brown, and Daniel Johnson

Faculty of Electrical Engineering and Computer Science,
Queensland University of Technology (QUT), Brisbane, Australia
joel.harman@hdr.qut.edu.au

Abstract. Eliciting accurate and complete knowledge from individuals is a non-trivial challenge. In this paper, we present the evaluation of a virtual-world based approach, informed by situated cognition theory, which aims to assist with knowledge elicitation. In this approach, we place users into 3D virtual worlds which represent real-world locations and ask users to describe information related to tasks completed in those locations. Through an empirical A/B evaluation of 62 users, we explore the differences in recall ability and behaviour of those viewing the virtual world via a virtual reality headset and those viewing the virtual world on a monitor. Previous results suggest that the use of a virtual reality headset was able to meaningfully improve memory recall ability within the given scenario. In this study, we adjust experiment protocol to explore the potential confounds of time taken and tool usability. After controlling for these possible confounds, we once again found that those given a virtual reality headset were able to recall more information about the given task than those viewing the virtual world on a monitor.

1 Introduction

The ability to accurately gather knowledge is a non-trivial challenge with relevance to a variety of domains [1]. Furthermore, if this information is not gathered effectively, there is the potential for many complications to arise. For example, when attempting to formalize business practices, working with incorrect or incomplete information may lead to higher construction costs, longer development times or poor-quality products [2].

Many approaches have previously been explored which aim to elicit information from stakeholders [3]. These methods, however, usually produce a trade-off between the quality of information gathered and the time required to do so [1]. For example, questionnaires can usually be administered easily, but often yield inaccurate or incomplete information. Role-plays on the other hand, have been shown to be an effective means of gathering accurate information [4], but are used less often in a business setting [3]. It has been suggested that this may be due to issues related to the potentially high setup costs and co-location issues related to role-play [1]. Despite these issues, in this paper we will be considering

a modified role-play approach due to the benefits it affords when looking to elicit accurate and complete information.

When performing a role-play, the individual steps in a task are usually executed in a specific sequence. The concept of memory chaining proposes that we are able to recall a piece of information more effectively if we are already thinking about the task it precedes [5]. This would suggest that even potentially mundane tasks, which may be ignored in other elicitation approaches (e.g. interviews), may initiate memory chains which could potentially yield relevant information. Furthermore, the rich context afforded during an in-situ role-play may also be sufficient in achieving a situated cognition response. Situated cognition is the theory that all knowledge is, to some extent, tied to the situations and contexts in which it was learned [6].

In our proposed approach, we look to leverage many of the benefits of a standard role-play. Rather than conduct the role-play within the real world, however, the session is instead conducted within a virtual environment closely resembling the real world location. By doing this, we aim to achieve many of the elicitation advantages associated with role-play, while mitigating many of the common detractors. Specifically, this approach may be preferable when a real world role-play would incur high setup costs, excessive risks, or have issues with co-location. For example, when looking to role-play a complex task, it may be difficult to arrange a time when all parties involved can be in the same location concurrently, without disrupting existing work commitments.

This is the third in a series of related studies exploring the potential recall benefits associated with providing situated context via a virtual world role-play. In our first study [7], we found that when comparing recall ability of participants provided with the virtual world stimuli and participants that weren't, participants given the stimuli were able to recall more information. In this first study, we conjectured that the stimuli provided by the virtual world was able to assist the users in recalling more information. Furthermore, we conjectured that if a more immersive interface was used (e.g. a virtual reality headset), a larger effect may potentially be observed. We later tested this theory in the second study and compared the recall ability of participants provided with a virtual reality headset with participants provided with a desktop display (monitor) [8]. The results of this experiment found that participants given the virtual reality headset were able to recall more information, but the experiment did identify some potential confounds which may have affected our results. In particular, we found that there were significant differences in both time taken and usability. Without further exploring these issues, we were unable to say with any degree of certainty whether the observed recall results were meaningfully due to the difference in viewing mechanism immersion, or some other secondary effect.

In the study presented in this paper, we aim to further explore this phenomenon and clarify the potential effect of confounds we found in our prior studies. In particular, we believe that the virtual world design and experiment setup we chose to use in our prior studies may have been too difficult for novice users, leading to issues with usability and subsequently affecting the time taken by

participants experiencing these issues. As the participants given the virtual reality headset were using a new interface, we believe that this issue was more pronounced within this group, leading to the differences in usability and time taken between the two groups. In this study, we look to obtain insight into these issues by making the task easier to learn, understand and execute for participants.

2 Situated Cognition and Virtual Worlds

Applying the concept of situated cognition within virtual worlds has been extensively explored in literature. When situated cognition was first proposed in 1989, it considered the concept of situated learning [9] to be a related concept [6]. Situated learning postulates that students need to be taught information within the contexts in which it will be applied.

Situated learning research has primarily considered virtual worlds for their potential applications to distance learning [10]. The goal of these environments is to place students within virtual classrooms, in order to better situate them within an environment they associate with a student-teacher learning dynamic. Some research even suggests that this approach may be an improvement over typical distance-learning communication methods, such as conference calls [11].

Outside of the context of learning, there has been an absence of literature which has explored the theory of situated cognition and the appropriateness of virtual worlds for achieving the necessary degree of situated context. We found one theoretical paper, however, which discusses the appropriateness of existing measures, such as immersion, and their importance in achieving effective situated cognition responses within virtual worlds [12]. Immersion, within the context of this study, is the degree to which a person's senses are engaged [13]. For example, a movie with sound may be considered more immersive than one without, as it engages the person's sense of hearing. As this study discusses the potential for assisting memory recall within virtual worlds via situated cognition, we have decided to primarily explore whether differences in immersion may affect user recall performance.

3 Research Questions

In this study, we have looked to explore how differences in viewing immersion may affect participant behaviour and memory recall ability. This had led to the formation of two distinct research questions:

- **RQ1:** How do changes in viewing immersion affect memory recall performance of users when they are asked to describe information while role-playing within a related virtual world?
- **RQ2:** How do changes in viewing immersion affect the behaviour of participants when they are asked to describe information while role-playing within a related virtual world?

We have chosen to explore potential behavioural differences as our prior studies have suggested that there may be a relationship between participant behaviour and recall performance [8]. To examine the effect of this change in immersion, we will be comparing a virtual reality setup, using an HMD, with a standard desktop display. From the two above research questions, we have constructed two main hypotheses after exploring existing theory within related literature. The first hypothesis is:

- **H1:** Users asked to recall information while viewing a related virtual world within an HMD will recall more information than those viewing the virtual world on a desktop display.

In prior work, we found preliminary evidence to suggest that the context provided by a virtual world was able to assist participants with recalling information [7]. We conjecture, however, that while the virtual world provided some benefit, it did not, wholly, afford the necessary context to the user. Specifically, we believe that the embodiment and immersion afforded to the user by virtual reality will better situate the user and assist them in recalling more effectively. In addition to recall improvements, we believe that the added embodiment and immersion provided by the HMD will also result in certain behavioural changes between the two groups. This has led to the formation of our second hypothesis:

- **H2:** Users provided with the HMD will be more exploratory in their approach than those provided with the desktop display.

To better evaluate this hypothesis, we will be exploring two sub-hypotheses. These are:

- **H2(a):** Users provided with the HMD will traverse a larger portion of the virtual world than those provided with the desktop display.
- **H2(b):** Users provided with the HMD will adjust their view within the virtual world more often than those provided with the desktop display.

While we do not necessarily believe that virtual reality innately makes a participant more exploratory, we do believe that the interaction mechanisms it affords makes it easier for participants to operate in a way in which they are familiar and experienced. While this hypothesis was somewhat supported in our prior experiments [7], we look to confirm that these effects remain present if the user does not need to continually halt movement to enter descriptions of their actions. Furthermore, we believe the evidence for increases in emotions [14], telepresence and sexual presence [15] when using an HMD, rather than a desktop display, are all examples of how users appear to behave more naturally within virtual reality. We wish to understand if this more natural behaviour may lead to observable increases in recall ability for the subjects.

4 Artefact Design

Virtual worlds are synthetic environments which provide users with an avatar through which they can explore and interact with other users, or the environment itself, to perform various tasks [16]. For this study, we have constructed a virtual world which aims to be representative of a real world airport.

A common issue related to research within this area, however, is that the virtual worlds can be difficult for participants to use when they are unfamiliar with the features and have not been provided with extensive training within the environment [17]. To mitigate this issue, we have developed a virtual world specifically to explore this recall phenomenon.

The virtual world was developed using the Unity3D [18] game engine. The environment was constructed using a mix of both constructed and prebuilt assets available via the Unity 3D asset store. Furthermore, we have also looked at trying to improve the usability of the virtual world in previous experiments [19] and made modifications to improve usability where possible. We have endeavoured to reconstruct the airport as best as possible. For this reason, we have included both critical areas (check-in, security, boarding and the plane itself), and non-critical areas (parking, shops, luggage collection and restrooms). Figure 1 shows screen captures of the developed virtual world.



Fig. 1. Screen captures of the developed virtual world airport.

To navigate the virtual world, participants from both the HMD and desktop display conditions were both given an Xbox 360 game controller. This controller had two main functions:

- The left joystick controlled avatar movement.
- The right joystick controlled avatar view changes.

We chose to use a game controller for this study as it allowed both treatment groups to use similar mechanisms for both navigating and viewing the environment. It must be noted, however, that as a core component of the HMD is to assist with adjusting avatar view, the HMD was still responsible for view changes within this condition and the vertical view component of view change was disabled on the controller joystick. Furthermore, both treatment groups

moved at the same constant speed and had the same turning speed when using the controller.

Unlike all prior experiments we have conducted to explore this phenomenon, the virtual world which was used in this experiment did not afford any interaction mechanisms to the user. Previously, participants were required to first look at an object or person within the virtual world, press a button to select the object and then write down the actions that they would normally complete with the object. While this approach greatly assisted in structuring the user responses, this did also present issues regarding usability. In this experiment, participants instead dictated all actions they would usually complete as they traversed through the virtual world.

5 Experiment Procedure and Measures

Participants who agreed to participate in the study were invited into a lab. They were then either presented with a 24 in. desktop monitor, or an HMD (Oculus Rift DK2), determined by random assignment. In addition to this, they were also provided with a game controller. A microphone was placed on the desk to record everything the participant said throughout the experiment. A between-subject design was chosen as we believed that participation in one treatment would meaningfully affect participant performance in the subsequent treatment. Upon entering the lab, participants read and signed the associated consent form and answered a series of control questions (e.g., knowledge of the boarding process, prior experience with virtual worlds). Following this, participants were placed into a virtual building, where they were trained to both view and traverse the environment. Participants were then instructed that for the experiment, they would be role-playing the task of boarding an airplane within a virtual airport. While doing this, participants were asked to verbally describe all the actions they would normally take within the environment, being as detailed and specific as possible. Participants were then placed into the airport to begin the experiment. No explicit time limit was placed on participants during their role-play. After participants were happy with their description of the task, they were asked to complete questionnaires related to presence and tool usability.

5.1 Recall Measure

The recall measure is the main item of interest in this study. To measure recall, we will be exploring the number of tasks described by participants within the two groups. This task count was generated by transcribing the logs from each session and manually assigning the phrases used by participants into individual tasks. For example, if a participant said “I hand my check in information to the boarding staff”, it would be counted as a single task. If the participant had instead specified “I hand my check-in information and luggage to the boarding staff”, however, this would instead be recorded as two different tasks (the giving of the check-in information and the giving of the luggage). We chose to use this

approach as we emphasized to participants during the session that they should aim to be as specific as possible when describing the actions that they would need to take.

To provide further insight into this result, we will also be comparing the total number of words used by participants in the two treatment groups. While this is not necessarily a measure of recall performance, it does provide an objective measure of how much the participants in both groups spoke. This may provide some validation for our recall measure if the differences between the two groups are similar (e.g. Group A described more than Group B because they spoke more), or create interesting insights if the number of words spoken result is not reflected in recall performance (e.g. Why did Group A describe less than Group B, despite talking more? What were they saying?).

To ensure that both treatment groups had a similar understanding of the task, participants were asked to both state the number of times they had been on a plane in the last five years and to rate their subjective knowledge of the airplane boarding task. As we did not prime participants with a description of the task prior to the experiment, and each participant may have their own nuanced personal approach for completing the task in the real world, we have opted to not examine the results for correctness. We chose not to prime participants as it provides a much more ecologically sound basis for the study. If we had primed participants, it would not be measuring long-term memory ability, but instead short-term memory of the description they were given. We have, however, removed tasks from our first measure where the statement was immediately identified as erroneous (e.g. the participant described the same task several times in a row or backtracked to try and correct a mistake). Finally, it should be noted that this task, while exploring recall ability, is technically measuring task performance. We believe that this is valid, however, as memory tests often require the participants to constrain the information they recall to the requirements of a task (such as a quiz) (e.g. [20,21]).

5.2 Exploration

We have chosen to examine how the users traversed the environment as we believe that this may provide insight into the behavioural approach taken by the participants. This is important as we believe differences in behaviour may influence recall performance. In our prior study, we found that participants given the HMD treatment explored a larger portion of the environment than those given the desktop display [8]. To measure this, we once again split the virtual environment into 216 5 m × 5 m segments (virtual world units were designed to be approximately equivalent to real-world units). For this analysis, we compare the average number of segments traversed by the two groups.

In our prior study, we found that participants given the HMD traversed more of the environment than those given the desktop display. As the way the two groups traversed the environment did not fundamentally change in this study, we expect a similar result will be observed.

5.3 Change of View

We will be exploring change of view factors in order to better understand how the two treatment groups examined the virtual environment. This paper was primarily motivated by the concept that inserting a user into a situation with the appropriate context would assist them in recalling information related to that situation. For this to be an effective approach, however, the user must view and understand this context. For this reason, we conjecture that, to some degree, viewing more of the environment may result in participants recalling more information about the task.

For this study, participants given the desktop display adjusted their view entirely with the provided game controller, while participants given the HMD adjusted their view with both the controller and the HMD. We will be considering view change as the average change of vertical viewing angle per minute. We have chosen to use only the vertical component of view change as the participants, in addition to making horizontal view changes to examine the environment, also adjust their view horizontally when they need to turn. This means that horizontal view changes do not necessarily indicate that a user is looking to adjust their view of their avatar specifically to view a different part of the environment. As vertical view changes were exclusively performed in order to explore the environment, we believe that this is the more accurate measure in this scenario. To measure these view changes in both groups, we will be examining the rotation made by the avatar, rather than the raw HMD tracking data. We have chosen to do this as we wanted to keep this measure consistent between the two treatment groups. Furthermore, the functions controlling avatar jitter automatically removed HMD jitter, and should therefore more accurately reflect the intentional view changes made by the users.

In our prior study, we found that participants given the HMD adjusted their view more often than those given the desktop display. As the way the two groups view the environment did not change in this study, we expect to see a similar result in this study.

5.4 Time Taken

This time taken measure considers only the time taken from when the participant begins describing the task to when they finalise their description. The time taken for all other activities (e.g. questionnaire completion) is not included in this measure. We have chosen to examine the time taken by both groups in the virtual world as its relationship has remained uncertain in our prior experiments. One existing study found that participants who described tasks faster tended to perform better [22], but their approach did not provide the situated context central to the approach presented in this paper. In our prior work, we found that participants given the HMD both took longer to complete the task and performed better. As we found no existing literature which explored a potential link between the time spent viewing stimuli and recall performance, we were unable to adequately determine whether the differences in recall performance

were due to differences in time taken between the two treatment groups, or due to the levels of immersion provided by the associated interfaces.

In our prior paper, we present several reasons why participants given the HMD may have taken longer to complete the task. For example, we identified that usability issues identified by participants given the HMD may have accounted for this difference in time taken. The design of this study aims to eliminate the differences in usability we found in our previous study by allowing participants to describe information verbally, rather than interacting with the world itself. For this reason, we will once again be examining time taken to determine whether there remains a difference in time taken by the two groups, despite the differences in experiment design.

5.5 Presence

Presence, as related to immersive virtual reality, is considered to be the concept of transportation. People are considered present when they report a sensation of being, to some degree, in the virtual world (e.g., you are there) [23]. We have chosen to measure presence in this study as existing literature suggests there may be a link between presence and recall performance [12]. In this study, we will be administering the Witmer and Singer presence questionnaire [24]. This questionnaire was chosen despite some criticism regarding its efficacy [25], as it remains the most widely administered survey for measuring presence in virtual worlds designed for non-game related activities.

In our prior study, we found inconclusive presence findings, with our quantitative results suggesting the HMD condition experienced lower presence while our qualitative findings suggested they experienced higher presence. We conjectured that this difference was due to the complexity of the task. Presence studies usually involve very simple tasks, where usability is not likely to be an issue. As this study aims to provide a much easier experience for the user to comprehend, we hope that the presence findings obtained in this experiment will not suffer from the inconsistencies encountered in our prior study.

5.6 Usability

We have chosen to measure usability as HMDs have been known to suffer from challenges with usability among novice users [26]. If a user has difficulty using the virtual environment, it may have ramifications on recall performance in addition to other aspects of the session. We have measured usability with the IBM usability satisfaction survey [27] as it remains a widely used measure for usability of software systems.

Furthermore, we are particularly interested in examining whether there are any usability differences between the two treatment groups as we found the HMD condition to have considerably lower usability in our previous study [7]. In this experiment, we have chosen to use a much simpler system for gathering information from participants, which we believe may mitigate this difference in usability between the two groups.

6 Results

Participants in this study were randomly assigned to either the HMD or desktop display conditions, with each condition having 31 participants. For comparing between the two conditions, we will be using two-tail Students-t tests when the data is both continuous and passes the Shapiro-Wilk W normality test ($p < 0.05$). If either of these conditions is not met, the two-tail Mann-Whitney U test will be used instead.

The average age of the participants was 24.84 (SD = 5.02). No significant difference was found between the age of those given the HMD and those given the desktop display ($p = 0.39$). Perceived understanding of the airport boarding scenario was quite high, with an average response of 5.40 (SD = 1.63) on a 7-point Likert scale. No statistically significant difference was found in perceived understanding between those given the HMD and those given the desktop display ($p = 0.41$).

Given the exploratory nature of the research presented and the relatively small sample size, we have not applied a Type I error correction (e.g., Bonferroni) to our analysis. Instead, we have elected to provide effect size calculations (Cohen's d) for all readings to provide the reader with a sense of the magnitude of all presented findings. Following Cohen (1992) [28], we treat all effect sizes between 0.2 and 0.5 as small, 0.5 and 0.8 as medium and greater than 0.8 as large. Due to the increased possibility of a Type I error, the findings presented in this study should be interpreted with a greater degree of caution. For consistency, we have used existing conversion formulas to convert the r effect sizes calculated for the Mann-Whitney U tests into comparable Cohen's d values [29].

6.1 Recall

In this study, we were primarily interested in comparing the recall ability of participants presented with the HMD and the desktop display. To evaluate H1, we compared the average number of tasks specified by both treatment groups. This test found that participants given the HMD were able to recall a larger number of tasks ($M = 10.50$, $SD = 2.48$) than those given the desktop display ($M = 8.97$, $SD = 2.72$, $U = 327$, $z = 2.15$, $p < 0.05$, $d = 0.57$).

To further explore this result, we also compared the average number of words spoken by both treatment groups. This test found that participants given the HMD spoke a larger number of words ($M = 212$, $SD = 47.3$) than those given the desktop display ($M = 189$, $SD = 20.2$, $U = 272$, $z = 2.93$, $p < 0.005$, $d = 0.80$).

6.2 Exploration

To evaluate Hypothesis 2(a), we have compared the mean amount of the environment traversed by the two groups. Results from this test found that participants given the HMD traversed a larger number of segments ($M = 88.2$, $SD = 27.4$) than participants given the desktop display ($M = 66.0$, $SD = 20.9$, $t(60) = 3.58$, $p < 0.001$, $d = 0.91$).

6.3 Change of View

To evaluate Hypothesis 2(b), we have measured the average amount participants adjusted the vertical view of their avatar. Results from this test found that participants given the HMD adjusted the vertical view of their avatar more often per minute ($M = 755$, $SD = 404$) than those given the desktop display ($M = 198$, $SD = 71$, $t(60) = 7.56$, $p < 0.0001$, $d = 1.92$).

6.4 Presence

Participants reported higher presence in the HMD condition ($M = 5.18$, $SD = 0.42$) than the desktop display condition ($M = 4.83$, $SD = 0.53$, $U = 282$, $z = 2.79$, $p < 0.05$, $d = 0.76$).

6.5 Usability

No statistically significant difference in usability was found between participants given the HMD and participants given the desktop display ($p = 0.16$).

6.6 Time Taken

No statistically significant difference in time taken was found between participants given the HMD and participants given the desktop display ($p = 0.09$).

7 Discussion of Results

The primary aim of this study was to further confirm whether improvements to interface immersion (specifically viewing immersion) may directly translate into better long term episodic memory recall. This work has been motivated by the continuing need to develop elicitation techniques which can effectively gather knowledge from individuals. Accurately gathering knowledge is a core component of many domains [1].

The results of this experiment supported our first hypothesis that participants given the HMD would be able to recall more information about the given scenario than those given the desktop display. It must be noted, however, that when asking participants to recall information, it is possible that we may have been exposed to issues regarding cognitive load. Cognitive load refers to the total amount of mental effort being used with regards to working memory. For example, if a task was to greatly tax working memory capability, it would be considered high cognitive load. It is possible that while participants were experiencing higher cognitive load at more complex parts of the task, they may have become less focused on accurately articulating their thoughts verbally. This may have inadvertently lowered the recall scores of participants in both treatment groups, as they may not have been talking as much during the more complex parts of their description. This is somewhat supported, as participants in our previous study did tend to recall more information when asked to write down their

knowledge of the task in a more structured manner. As both treatment groups were asked to describe the same task, however, the two treatments should have required similar levels of cognitive load. For this reason, we do not believe this finding invalidates the observed results. While the observed recall findings do suggest that the immersive system of virtual reality may be able to assist with allowing users to better recall information, we have also explored several other items of interest to better explain the observed result.

In addition to looking at memory performance, we have also explored behavioural differences between the two groups to determine whether any differences in behaviour may have also potentially affected recall performance. When examining how the two groups traversed the environment, we found that participants given the HMD explored more of the environment than those given the desktop display (supported by a large effect size). When examining how the two groups viewed the environment, we found that participants given the HMD adjusted their view more often than those given the desktop display. As participants given the HMD both traversed more of the environment and viewed more of the environment, we believe these findings support H2(b) and indicate that the participants given the HMD were more exploratory in their role-play approach. While it is likely this difference in view change was due to the ease at which the participants given the HMD were able to adjust their view to examine the environment, we do not believe that this diminishes the significance of the finding. The ability to easily change view is a fundamental affordance provided by the HMD. While we cannot say, with any degree of certainty, that these differences did result in better recall performance, these findings do seem to be consistent with existing literature which suggests that providing participants with adequate levels of context is important to best achieve a situated cognition effect [6].

No statistically significant difference in usability scores was found between those given the HMD and those given the desktop display. While this result is in no way conclusive, this suggests that the possible link between time taken and recall performance may not be as significant as our previous experiment suggested. This is a positive result, as one of the main reasons for running this study was to eliminate, or at least partially mitigate, the difference in usability between the two groups. Ensuring similar usability between the two treatments reduces the chance that usability differences may have affected any of the other results presented in this study.

In addition to usability, we also looked to explore any potential difference in time taken between the two groups, as differences in time taken presented a potential confound in our prior studies. In this experiment, we found no statistically significant difference in time taken between the HMD and desktop display conditions.

Finally, participants given the HMD reported higher subjective presence scores than those given the desktop display. These findings are consistent with prior research [23], which discussed a possible link between presence and memory performance within virtual worlds. Due to the increased levels of presence experienced by those given the HMD, we conjecture that the higher presence

provided by the immersive properties of the HMD interface may be responsible for the improved recall ability. Further work will be required to better explore this phenomenon and adequately identify the exact mechanism, or mechanisms, facilitating better recall while immersed within virtual reality.

8 Synthesis from Overall Results

In this section, we will synthesise the findings discussed above with the data gathered in our prior studies. By doing this, we aim to identify patterns across different studies which may provide further insight into the overall approach. Specifically, we will be looking at the outcomes of this study and two prior studies to identify potential patterns and discuss possible reasons for differences in results across each of these studies. To assist with this, Table 1 provides a brief overview of the three experiments as well as a brief summary of the pertinent results. As we will be discussing multiple studies in this section, we will refer to the studies as they are numbered in Table 1, referring to the initial study we conducted as *Study One*, our prior study which also investigated recall differences between HMDs and desktop displays as *Study Two* and the study which was presented in this paper as *Study Three*.

Study One provided the initial motivation for the approach which was subsequently explored in Study Two and Study Three. Study One provided evidence that suggested the context provided by virtual worlds may assist in improving recall ability. Due to the differences in experiment design between the first study and the other two studies (e.g. differences in the virtual world visuals), we will not be making direct comparisons between the results of the first study and the results in the other two studies. We did feel that including this study in Table 1 was warranted, however, as it provided the initial results which assisted in designing the two subsequent studies. As both Study Two and Study Three had many similarities, however, we will be primarily comparing the results of these two studies and discussing potential causes for any differences in observed results.

In both Study Two and Study Three, we found that participants given the HMD were able to recall more information about the airport boarding task than those given the desktop display. This provides further evidence that the immersive qualities of the HMD were able to improve memory recall capabilities of participants.

Furthermore, in both Study Two and Study Three we found that participants given the HMD were also more exploratory in their approach, choosing to both traverse more of the environment and look at more of the environment. In Study Two, however, we found significant differences in the traversal patterns of the two groups. This was not the case in Study Three, whereby both groups traversed all areas of the airport, despite the HMD group traversing more on average. Furthermore, we found that both treatment groups traversed more of the environment than in Study Two. This may be due to the removal of the user interface, which forced users to stop moving and enter information into the system intermittently.

Table 1. Description and results summary for all related studies.

Study 1: No Context Compared with Virtual World Context [7]
This first study compared the recall ability of participants provided with a virtual world (viewed on a desktop display) with participants provided with a 2D process description environment. This study found: <ul style="list-style-type: none"> - Participants given the virtual world recalled more information - Participants given the virtual world completed the task faster
Study 2: HMD Compared with Desktop Display (Interactions) [19]
This second study compared the recall ability of participants provided with a virtual world viewed on either an HMD or a desktop display (monitor). Participants described information by interacting with objects and writing descriptions. This study found: <ul style="list-style-type: none"> - Participants given the HMD recalled more information - Participants given the HMD viewed and traversed more of the environment - Participants given the desktop display completed the task faster - Participants given the HMD reported lower subjective usability scores - Participants given the HMD reported lower subjective presence scores
Study 3: HMD Compared with Desktop Display (No Interactions)
The third study also compared the recall ability of participants provided with a virtual world viewed on either an HMD or a desktop display (monitor). Unlike in the second study, however, the participants could not interact with objects in the virtual world. Instead, users dictated the actions they would usually perform with each of the items or people as they traversed the environment. This study found: <ul style="list-style-type: none"> - Participants given the HMD recalled more information - Participants given the HMD viewed and traversed more of the environment - No significant difference was found in time taken between the two groups - No significant difference was found in usability between the two groups - Participants given the HMD reported higher subjective presence scores

In Study Three, we found no statistically significant differences in usability between the HMD and desktop display conditions. While not conclusive, this is a positive result, as the differences in usability identified previously in Study Two put many of our observed results into question. Issues with usability can manifest in a variety of ways and may have affected many, if not all of our results in some way. By simplifying the experimental setup and user requirements for this experiment, we have greatly mitigated any potential confounds posed by usability.

In Study Three, we also found no statistically significant difference in time taken between the HMD and desktop display conditions. Once again, this result is contrary to what we found previously in Study Two. This difference in time taken in our prior study was problematic, as we were unable to dismiss the possibility that the observed recall differences were not due to the differences in time

viewing the given virtual world stimuli. This meant that it may not have been the immersive qualities of virtual reality producing differences, as we discussed in our research questions and hypothesis construction. In our prior experiment, we conjectured that a major reason for the difference in time taken between the two groups was due to the significantly lower usability scores reported by those given the HMD. If the participants given the HMD were finding the virtual world harder to use, it may mean that they would require longer to complete the task. This suggests that the observed difference in recall ability between the two treatment groups was likely a direct result of the way the two groups viewed the environment, rather than simply the time they spent viewing the virtual world stimuli.

Finally, Study Three found that participants given the HMD reported higher subjective presence scores than those given the desktop display. In Study Two, our presence findings were inconclusive. The subjective presence questionnaire suggested people given the HMD experienced lower levels of presence, while the semi-structured interviews suggested that they were experiencing higher levels of presence. After reflecting on the usage of the presence survey, however, we believe that the task we provided to participants in Study Two was not well suited to this particular questionnaire. After reviewing many of the papers which use this particular survey, the tasks given to participants tended to be very simple and contained a very unobtrusive user interface. In Study Two, however, the participants had to use a variety of menus and object interactions to describe their desired actions. We believe that these menus may have inadvertently affected the user presence scores on the chosen survey. Despite this, however, we have no way of knowing for certain which treatment group experienced greater levels of presence in Study Two. In Study Three, however, we have removed all menus from the task and the HMD subjective presence scores are now also significantly higher than the desktop display scores.

We believe that the study described in this paper has meaningfully clarified and explored the issues identified in our prior studies. Furthermore, this work has provided further rigor to our prior results and provided further evidence of a link between the added immersion provided by virtual reality and improved recall performance.

9 Design Insights

As we have now constructed multiple virtual worlds with varying interfaces, we believe it is prudent to discuss some of the insights we have gathered during the development, testing and refinement of these environments. While there are numerous design decisions made when looking to develop a virtual environment for eliciting information, our work in this area has generated particularly beneficial insights regarding how the virtual environments should be constructed and how the information should be elicited from the user.

Despite significant advancements to the tools and resources available for constructing these environments, the time required to accurately reconstruct

environments manually is still greatly prohibitive to the overall approach. The environment used for this experiment was constructed in a matter of hours, but further time may be required for more complex environments, especially if objects within the environment needed to be created. Unless the virtual world reconstruction had already been developed for other purposes (e.g. training simulations), it would likely still be easier to conduct real-world role-plays, rather than manually attempt to reconstruct these environments. These environments do not manually need to be constructed, however, and may instead be generated via 3D scanning techniques (e.g. from a Matterport depth camera [30]). In our experiments, we chose not to use this approach as we did not want user experience within the real-world airport we chose to scan to confound our results. We have worked with these scanned environments, however, and we believe that they provide a sufficient level of detail for this particular approach, while greatly reducing the time required to generate these environments.

After working with both a structured approach (where participants manually entered their descriptions of individual tasks) and an unstructured approach (where participants dictated their descriptions of tasks), we also believe that providing some structure to the way in which participants provide their descriptions of each task is beneficial. In the experiment described in this paper, we believe that there may have been times where participants were role-playing the tasks they wished to complete within the environment, but forgot that they also needed to also describe their actions aloud. While this may have also occurred in our previous experiments, we do not believe it was so pronounced. This may be because participants given the structured approach believed they were constructing a list of required tasks for later review, while those dictating their actions believed they were describing their actions forthrightly. This is somewhat supported as many participants which described their actions using the structured approach refined their descriptions over time, while those dictating their descriptions were much less likely to make revisions. For these reasons, we believe that a structured approach is preferable when attempting to elicit information. This structure, however, may not necessarily have to be achieved by having participants manually enter descriptions. For example, this may instead be achieved by having the system only record while the user holds a button down. When the user then releases the button, the system then stores the description and allow the user to review and modify all of their prior recordings. Specifically, we believe that the key difference between these two approaches is the ease in which participants can review and revise their previous descriptions.

While we believe that these two areas are important when looking to develop a virtual world for elicitation, considerable future work is required to adequately explore the many design challenges associated with constructing a virtual world for this purpose.

10 Conclusion and Future Work

In this study, we have continued to explore the potential of using a virtual world role-play approach in order to assist in effectively eliciting knowledge. This

approach was motivated by the theory of situated cognition, which postulates that by situating a user within a specific context, it becomes easier for them to recall related information. This study aimed to solidify the results of prior studies by exploring potential confounds related to usability, presence and time taken.

The results from this study indicated that participants using the HMD both described more tasks and spoke a larger number of words than those given the desktop display. When comparing the behaviour of the two groups, we also found that participants given the HMD also tended to traverse more of the environment and modify their view more often (supported by large effect sizes). Unlike in our prior study, no statistically significant difference in time taken or usability was found between the two treatment groups. In addition to this, the HMD condition reported higher presence than the desktop display condition. These results supported our prior conjecture that the interaction mechanisms presented by the virtual world in our previous experiment may have inadvertently affected the observed results.

In this experiment, however, we were measuring performance based on the dictation provided by participants. This does not necessarily adequately align with the actual recall ability of participants. There is a possibility that the areas of the task which would normally require high cognitive load may have resulted in participants actively describing less. Future work will be required to determine the effect that varying cognitive load may have on the ability of participants to adequately describe their knowledge. Furthermore, as both this study and prior studies used the same airport scenario, further research is required to adequately explore whether the observed results may generalise into other complex spatial scenarios, such as warehouse inventory management or hospital medical processes.

References

1. Davis, A., Dieste, O., Hickey, A., Juristo, N., Moreno, A.M.: Empirical results derived from a systematic review. In: *Proceedings of Requirements Engineering*, pp. 179–188. IEEE (2006)
2. Smith, E.A.: The role of tacit and explicit knowledge in the workplace. *J. Knowl. Manag.* **5**(4), 311–321 (2001)
3. Zowghi, D., Coulin, C.: Requirements elicitation: a survey of techniques, approaches, and tools. In: Aurum, A., Wohlin, C. (eds.) *Engineering and Managing Software Requirements*, pp. 19–46. Springer, Berlin (2005). doi:[10.1007/3-540-28244-0_2](https://doi.org/10.1007/3-540-28244-0_2)
4. Costain, G., McKenna, B.: Experiencing the elicitation of user requirements and recording them in use case diagrams through role-play. *J. Inf. Syst. Educ.* **22**(4), 367 (2011)
5. Mace, J.H., Clevinger, A.M., Martin, C.: Involuntary memory chaining versus event cueing: which is a better indicator of autobiographical memory organisation? *Memory* **18**(8), 845–854 (2010)
6. Brown, J., Collins, A., Duguid, P.: Situated cognition and the culture of learning. *Educ. Res.* **18**(1), 32–42 (1989)

7. Harman, J., Brown, R., Johnson, D., Rinderle-Ma, S., Kannengiesser, U.: Augmenting process elicitation with visual priming: an empirical exploration of user behaviour and modelling outcomes. *Inf. Syst.* **62**, 242–255 (2016)
8. Harman, J., Brown, R., Johnson, D.: The role of immersion during situated memory recall within virtual worlds. In: *Proceedings of the 28th Australian Conference on Computer-Human Interaction*, pp. 1–10. ACM (2016)
9. Lave, J., Wenger, E.: *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press, Cambridge (1991)
10. Dede, C., Nelson, B., Ketelhut, D.J., Clarke, J., Bowman, C.: Design-based research strategies for studying situated learning in a multi-user virtual environment. In: *Proceedings of Learning Sciences 2004*, pp. 158–165 (2004)
11. Dickey, M.D.: Teaching in 3D: pedagogical affordances and constraints of 3D virtual worlds for synchronous distance learning. *Distance Educ.* **24**(1), 105–121 (2003)
12. Carassa, A., Morganti, F., Tirassa, M.: A situated cognition perspective on presence. In: *Proceedings of Cognitive Science 2005*, Sheridan Printing, pp. 384–389 (2005)
13. Mania, K., Chalmers, A.: The effects of levels of immersion on memory and presence in virtual environments: a reality centered approach. *CyberPsychol. Behav.* **4**(2), 247–264 (2001)
14. Riva, G., Mantovani, F., Capideville, C.S., Preziosa, A., Morganti, F., Villani, D., Alcaiz, M., Gaggioli, A., Botella, C.: Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychol. Behav.* **10**(1), 45–56 (2007)
15. Renaud, P., Rouleau, J.L., Granger, L., Barsetti, I., Bouchard, S.: Measuring sexual preferences in virtual reality: a pilot study. *CyberPsychol. Behav.* **5**(1), 1–9 (2002)
16. Duncan, I., Miller, A., Jiang, S.: A taxonomy of virtual worlds usage in education. *British J. Educ. Technol.* **43**(6), 949–964 (2012)
17. Virvou, M., Katsionis, G.: On the usability and likeability of virtual reality games for education: the case of VR-ENGAGE. *Comput. Educ.* **50**(1), 154–178 (2008)
18. Unity Technologies: Unity - Game Engine (2016). Accessed 13 June 2016. <http://unity3d.com>
19. Harman, J., Brown, R., Kannengiesser, U., Meyer, N., Rothschädl, T.: Model as you do: engaging an S-BPM vendor on process modelling in 3D virtual worlds. In: Fleischmann, A., Schmidt, W., Stary, C. (eds.) *S-BPM in the Wild*, pp. 113–133. Springer, Cham (2015). doi:10.1007/978-3-319-17542-3_7
20. Roediger, H.L., Karpicke, J.D.: Test-enhanced learning taking memory tests improves long-term retention. *Psychol. Sci.* **17**(3), 249–255 (2006)
21. Ivanoiu, A., Adam, S., Van der Linden, M., Salmon, E., Juillerat, A.C., Mulligan, R., Seron, X.: Memory evaluation with a new cued recall test in patients with mild cognitive impairment and Alzheimers disease. *J. Neurol.* **252**(1), 47–55 (2005)
22. Claes, J., et al.: Tying process model quality to the modeling process: the impact of structuring, movement, and speed. In: Barros, A., Gal, A., Kindler, E. (eds.) *BPM 2012*. LNCS, vol. 7481, pp. 33–48. Springer, Heidelberg (2012). doi:10.1007/978-3-642-32885-5_3
23. Schumie, M.J., Van der Straaten, P., Krijn, M., Van der Mast, C.A.P.G.: Research on presence in VR: a survey. *Cyberpsychol Behav.* **4**(2), 183–202 (2001)
24. Witmer, B.G., Singer, M.J.: Measuring presence in virtual environments: a presence questionnaire. *Presence: Teleoperators Virtual Environ.* **7**(3), 225–240 (1998)
25. Schubert, T., Friedmann, F., Regenbrecht, H.: The experience of presence: factor analytic insights. *Presence* **10**(3), 266–281 (2001)
26. Sutcliffe, A.G., Kaur, K.D.: Evaluating the usability of virtual reality user interfaces. *Behav. Inf. Technol.* **19**(6), 415–426 (2000)

27. Lewis, J.R.: IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *Int. J. Hum.-Comput. Interac.* **7**(1), 57–78 (1995)
28. Cohen, J.: A power primer. *Psychol. Bull.* **112**(1), 155 (1992)
29. Rosenthal, R.: Parametric measures of effect size. In: Cooper, H., Hedges, L.V. (eds.) *The Handbook of Research Synthesis*, pp. 231–244. Russell Sage Foundation, New York (1994)
30. Matterport: Immersive Spaces for Real-world Applications - Matterport (2016). Accessed 23 June 2016. <http://matterport.com>