Virtual Reality as a Context for Memorizing: It Can Make a Difference

Graduation Thesis
Maik Lanen
Thesis advisors: Maarten H. Lamers and
Jasper Schelling (Rotterdam University of Applied Sciences)

Media Technology MSc Program, Leiden University, The Netherlands
June, 2018

In this research 51 participants of Dutch origin performed three memory tasks either in the context of virtual reality or reality. Twenty-four hours later they were tested under the same or different conditions. The main result found in this research is that the groups without a context change scored significantly higher on the overall score than the groups with the context change. The result imply that it is much more debatable if we can use learning in virtual reality on real life situations whereby the ability to recall plays a role. Another finding is that the context group virtual reality - reality scored much lower on the overall score than the rest of the groups. After a comparison between the virtual reality - reality group and the reality - virtual reality group, we found an almost statistical difference, which results in a non-statistical effect that stays unexplained.

KEYWORDS: *virtual reality, memory, context, context-dependent memory*

1. Introduction

In this paper, we will discuss the field of context-dependent memory and its relation to virtual reality. Though there has been a lot of research done in the past between memory, context, and virtual reality, we believe there is still some information to be explored and uncovered.

It is possible that you have had to memorize words for a secondary language. You are sitting in your room, for example, and you memorize the given selection words until you can recall them well. The next day at school, you are quizzed on your ability to recall the words and you catch yourself forgetting a portion of the words. Is it because you have started studying too late and did not give yourself enough time? Is it because you did not study hard enough? However, when you come home and you are sitting behind your desk again, somehow you can recall most of the words. How is that possible?

This is because there is a phenomenon that is called context-dependent memory whereby context influences the ability to recall items from your memory; this is a subject which has been researched by Baddeley in 1975. Baddeley let two groups learn words in two different contexts, namely under and above water, and let them recall the words in either the same context or a different context. He found out that the test persons who had to recall in a different context scored significantly worse than the test persons who had to recall in the same context as they have learned the words in. This study shows that the change in context negatively affects the ability to recall items learned in another context.

The relation of context-dependent memory and virtual reality is an interesting topic to research because the role of virtual reality has become more prevalent to our society. Virtual reality is already being used to help people with medical conditions like brain injuries and depressions. Our research could help to uncover knowledge about context-dependent memory and its relation to virtual reality. For example, think about a military rescue team who had learned the position of hostages in a virtual environment. Will the team remember where the hostages are in reality, or will the change in context from virtual reality to reality affect their memory? A less radical example could also explore the relation between virtual reality and education. Can a learning program in virtual reality be used in class to help people learn and memorize better? Will it affect their memory in a negative way, or perhaps not at all?

In this paper, we are interested to research if a change in contexts, namely virtual reality and reality, affects the ability to recall. Our research question is therefore: Does a change in context between virtual reality and reality influence the ability to recall? Our hypothesis is that the change in context between virtual reality and reality has an effect on the ability to recall.

We will firstly explore the most important related work done to-date, specifically in relation to memory and context, and memory and virtual reality. We will explain our method after the related work and talk more in detail about how our experiment was done and why we made the choices we had. After the method, we will discuss the results of the experiment and see if it led to some findings. Lastly, we will discuss the results, make conclusions, and determine if our hypothesis was correct.

2. Related Work

2.1 Memory and Context

We will begin by talking about the main research that has been done in the field of context memory. We will not discuss every research that has been done in the field of context memory, but only the studies that most apply to our research.

In 1969, Goodwin *et al.* researched the influence of context in relation to memory. The researchers performed an experiment whereby the participants had to perform four different memory tasks while being sober or being under the influence of alcohol. After twenty-four hours, the participants were to return and be tested under both the same conditions. The results showed that the participants were better in recalling when being tested in the same state they learned in. The results also showed that the participants' ability to recognize had not been altered by the different states; thus, not all subforms of memory are affected.

A study very close to this research had been performed by Godden and Baddeley in 1975 whereby divers had to learn lists of words in two contexts: underwater and on land. The study showed that when the divers had to recall the words in the same context as when they had learned it, the ability of recalling the words had been improved compared to trying to recall in a different context. The experiment shows that the disruption in moving from one environment to another influences the ability to recall memory learned in the first environment.

The context-dependent memory effect was also researched by Johnson and Miles in 2008 by investigating the influence of taste on memory. Johnson and Miles let two groups take a test whereby the participants had to learn and recall. One group had to take the test by chewing on flavourless gum and the other group had to take the test chewing on mint-flavoured strips. The result of the experiment is in contrast to the previous discussed studies. There was no context-dependent memory effect found with either flavourless gum or mint-flavoured strips, which indicates that the given context was not strong enough or insufficient to produce the effect.

In the study of Smith, Glenberg, and Bjork in 1978, the participants had to recall and recognize different elements, like lists of words and contextual information. The results showed that the test subjects performed better when there was a variability of input environments in contrast to an unchanged environment. The study also showed that recalling words within a category worked better for the same context than a different context recall, which is in line with the previously two discussed studies.

Another way of influencing the context of memory was researched by Bower, Monteiro, and Gilligan. The researchers looked at whether a person’s mood could act as a context for learning and recalling. In this study, the test subjects were to learn a word list in the state of being happy or sad and had to recall the list the same day and after twenty-four hours, and in the two different states. Just as the previous research would suggest, the same results occurred; the test subjects were significantly better at recalling the list in the same state they learned in, as opposed to recalling in the changed state.

2.2 Memory and Virtual Reality

We are now going to discuss virtual reality and memory, and explore what has been researched so far in this field. In July 1996, there was a study by Attree *et al.* that researched the potential value of virtual reality in relation to treat memory deterioration and how participation plays a role in it. The researchers investigated if there was an effect on the memory between active and passive participation in a virtual environment. It was found that the participants with active participation tested as having a better memory for spatial layouts and the participant who passively participated tested higher in object recall.

Another way in which memory was researched in combination with virtual reality was by looking at the possibility of enhancing someone’s memory by making virtual reality as realistic as possible for the user. The study from March 1999 by Dinh *et al.* researched this idea in which the participants underwent a virtual experience whereby they could smell, feel, and hear the virtual environment. It was predicted that the more realistic the virtual environment would be, the better the recall for memory would be. The results showed that, by tactile input, the quality of presence in virtual reality was being enhanced. The tactile input also made it easier for the participants to remember the objects in the virtual environment. Auditory and olfactory stimulants only increased the feeling of being present in the virtual world, but had no relation to memorizing better.

Another interesting study was done by Plancher *et al.* (2012), where they explored if a difference in age in combination with memorizing in virtual reality had an impact on memory. The researchers let all participants simulate driving with and without the intention to memorize their driving route. The result was that the group with the younger participants scored better in recalling the driven route in both scenarios than the older group.

A different way of studying memory in relation to virtual reality has been done by Gould *et al.* in 2007. The research focussed on the effects of spatial memory in depression, whereby virtual reality played a key role in assessing. The participants consisted of 30 depressed patients and 19 comparison subjects. The participants were asked to find locations in a virtual town. The study showed that depressed patients performed worse than the comparison group in the spatial memory assignment. The study from Gould *et al.* is not the first study to research the relation between depressed patients and spatial memory. According to Gould *et al.*, there have been studies that showed inconsistent results in regard to this subject. With this study, the researchers aimed to form consistent results in the area of depression and spatial memory by using virtual reality as a key tool.

Another study whereby virtual reality is used to assess memory in relation to a medical disease is the study from Sweeney *et al.* in 2010. In this study, the researchers looked at the executive memory of the participants with a non-progressive brain injury and reported executive difficulties in everyday life using virtual reality as an assessment tool. The participants were asked to work as a manager in a storage unit company in virtual reality. The participants had to execute tasks whereby planning, rule following, and prospective memory tasks are important. The result of the test was that the group with the brain injury scored significantly worse than the control group- specifically in their strategies, time-based prospective memory, and time to complete tasks.

2.3 Key Findings

2.3.1 Key findings

KF 1: The tasks given to the participants in the researches of Goodwin *et al.* and Marks *et al.* are found to be useful for our experiment. We will use the following tasks from the study in our experiment: the rote-learning task, the association task and the recognition task.

KF 2: In the research of Johnson and Miles in 2008 they gave the participants two minutes time to memorize a list of words.

KF 3: In the research of Godden and Baddeley they gave the participants two minutes to recall the learned items.

KF 4: The commonly used time between memorizing and recall is twenty-four hours in most studies.

KF 5: The results of the study from Dinh *et al.* shows that, by tactile input, the quality of presence in virtual reality was being enhanced.

KF 6: The results of the study from Dinh *et al. also* shows that, tactile input makes it easier for the participants to remember objects in the virtual environment.

KF 7: The results of the research of Plancher *et al.* shows that age has an influence on memorizing in virtual reality. The group of younger participants tend to score better than the older participants by recalling in virtual reality.

2.3.2 Summary

We see that there has been a lot of research done on memory and its relation to context and memory in virtual reality. If we look at the spectrum of these three topics in Figure 2.1, we can see that there has been research done in memory and its relation with context, memory and its relation to virtual reality, and virtual reality in relation to context. So far, there has not been research done regarding the relation between all three topics. In this paper, we will look if the change in context by experiencing virtual reality versus reality affects the ability to recall memory.

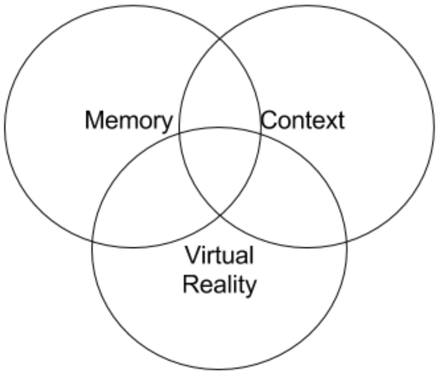


Figure 2.1: Venn-diagram depicting the areas of interest for our study

3. Method

In the study, we have four groups of participants who need to learn items in relation to memory recall. Every group is going to learn in a context and recall in the same or a different context. You can see the deviation of the groups in Figure 3.1. For example, you can see the underlined text in Figure 3.1, if a participant memorized the items in the virtual state, they had to recall in reality. After 24 hours, the participants will be assessed on their ability to recall the items [KF4]. In practice, assessing all the participants after exactly 24 hours will be difficult, thus we aim to assess after 24 hours and keep exact times of memorizing and assessing. We use this data to check if it has influenced the results of the experiment.

	Learned in VR	Learned in reality
Recalled in VR	participants	participants
Recalled in reality	participants	participants

Figure 3.1: Table showing the distribution of the participants

3.1 Environment

We will firstly talk about the environment of the virtual reality test. The test person will sit on an office chair behind a desk with the virtual reality headset on (see Figure 3.2). The virtual reality headset that we will use is the HTC Vive. We choose the HTC Vive because this headset has two joysticks which can sense motion and make interaction in virtual reality possible for the user.

Figure 3.2: Test person experiencing the virtual reality context in the attic of Bever

In the virtual world, the participants will sit behind a desk with a book in front of them and he or she will be surrounded with divider boards in a U-shape (see Figure 3.3). We choose to work with the divider boards because it allows us to test in multiple locations in reality without changing the environment in the virtual reality. There will also be a virtual plant on the table and a virtual poster on the divider board, just like in the real world [KF5]. We added these features so the context would be more enhanced and so there would be more for the test person to experience.

Figure 3.3: Overview of the virtual reality setup in Unity

In the virtual environment, the test person is able to look around and interact with the book on the table. The interaction consists of the ability to turn over a page with the use of the joysticks from the HTC Vive set. The interaction is as follows: the test person moves his virtual hand towards the top right corner of the book, clicks and holds the trigger on the controller and makes an overturning movement to thereafter release the trigger. In virtual reality, the test persons are also able to see their own avatar’s arm. We choose to give this arm a blue sleeve and a light toned skin color because most of the test persons that we will test will have a light skin tone and may be able to better identify themselves with the hand.

Though the environment may have changed, efforts have been made to ensure reality would look as similar as possible to the virtual reality environment explained before. We used the same type of plant and office chair as in the virtual world and made the divider with the poster as similar as possible (see Figure 3.4). The test person is asked to have the same interaction with the environment as in the virtual world. The testers, just like in virtual reality, will be only allowed to look around and turn over the page of the book in front of them. The experiment will be executed in two different locations: namely, in the attic of a Dutch outdoor sports retailer called Bever, and in a designated testing location at the Rotterdam University of Applied Sciences.

Figure 3.4: Overview of the test setup in the attic of Bever

3.2 Demographic Questions

Before the demographic questions, the participant will be verbally informed about the procedure of the test whereby all parts of the test will be explained in detail. The test person will be given two language options in which he or she can perform best in the experiment. The first option is in Dutch and the other option is in English. We provide the experiment in these two languages because during the pilot, we saw that some Dutch participants were having difficulties performing the experiment in English. We also think that the lack of fluency in a language could affect the results of the experiment; thus a participant can only participate the experiment when he or she is fluent in Dutch or English. The decision if somebody is fluent enough in a language will have to come from the participants themselves. In the experiment, we will make sure that the distribution of English and Dutch participants will be the same per group. For example, if we have in total forty participants of which thirty two Dutch speaking and eight English speaking, then the distribution of the Dutch and English speaking participants must be the same per group as shown in Figure 3.5.

	Learned in VR	**Learned in reality**
Recalled in VR	8 Dutch, 2 English	8 Dutch, 2 English
Recalled in reality	8 Dutch, 2 English	8 Dutch, 2 English

Figure 3.5: Table showing an example of the distribution of Dutch and English speaking participants

After the introduction, the supervisor will ask the participant general questions like their sex, age [KF7], experience with virtual reality, and education. The answers will be processed in an Excel sheet. The demographic questions are formulated by the standards of the PGA Group [15] and the questions about education are formulated by the European Qualifications

Framework [14]. These demographic questions are included in the experiment to gather general information related to the test about the participants. The data will be assessed to see if, for example, age or education of the test person has an influence on the results.

3.3 Orientation Task

After the demographic questions, the participants will be introduced to their given contexts: namely, virtual reality or reality. The participant will be verbally asked to execute a simple orientation task in the given context. This task is made to make the participants familiar with the given context and to make sure the test person is proficient enough technologically to participate with the experiment. The test consists of 1) looking around at the different objects and 2) executing an example task whereby the participant has to interact with the book on either the real or virtual table.

The test person was asked firstly to look at the plant, then the poster, and lastly the book. This meant looking left, right and down from their position. After the orientation, the test person was asked to execute an example task. The example task consists out of interacting with the book by turning from one page to the next.

On the day of recall, the participants were asked to do the same orientation task as on the day of memorizing. We let the participants do the orientation task again in case they had forgotten how the procedure works and so they can get used to the given context again.

3.3.1 Assessment

The orientation task is considered a success if the test person is able to look at the objects and successfully interact with the book. If the participant is not able to perform the given interactions then the supervisor will try to help the participant to master the interactions. If the participant is not able after the help from the supervisor to master the interactions, then the test person will be excluded from the experiment.

3.4 Rote-learning Task

After the introduction, the participants will complete different tasks which are based on the studies from Goodwin *et al.* in 1969 and Marks *et al* in 1964 [KF1]. These tasks are the rote-learning task, the association task and the recognition task. The first test is a verbal rote-learning task to measure the ability to recall.

3.4.1 Day 1: Memorizing

On the day of memorizing, the participants are asked to perform a rote-learning task. The rote-learning task consists of saying four 5-word sentences with varying meaningfulness, out loud, repeatedly. The amount of time given for this task is two minutes. The amount of time is based on a similar study, namely the study from Johnson and Miles in 2008 [KF2]. The test persons will be asked to remember the sentences and they will be assessed on the ability to recall after twenty four hours. There are four types of sentences: a normal sentence, an anomalous sentence, an anagram, and a word list. The normal sentence was gathered from a magazine. The selection of the sentence was partly random, the only requirement was that the sentence had to be five words long. The anomalous sentence used in this task is fetched from the paper called “Three models for the description of language” by Chomsky, whereby he discusses what an anomalous sentence is [13]. The English five word anagram is generated by an online anagram generator provided by litscape.com [16].

The random words are generated by an online tool called Textfixer [17]. All the sentences used in this task are translated from English to Dutch, except for the anagram. The Dutch anagram is provided by mijnwoordenboek.nl [18]. For the English participants the word list is in English (see Figure 3.6) and for the Dutch participants there is a Dutch word list (see Figure 3.7). The sentences are provided in the book for the reality group and in the virtual book for the virtual reality group.

- I walk to the station
- Colourless green ideas sleep furiously
- Drawer, Redraw, Reward, Warder, Warred
- Flatness, Iron, Harbor, Crab, Thief

Figure 3.6: Four English five-word sentences used in the rote-learning task in the order of normal sentence, anomalous sentence, anagram and word list

- Ik loop naar het station
- Kleurloze groene ideeën slapen woedend
- Mentors, Stormen, Stromen, \`n Stomer, \`t Morsen
- Vlakheid, IJzer, Haven, Krab, Dief

Figure 3.7: Four Dutch five-word sentences used in the rote-learning task in the order of normal sentence, anomalous sentence, anagram and word list

3.4.2 Day 2: Recall

On the day of recall the participants are asked in the given context to recall the sentences learned twenty four hours prior. The participants have two minutes to recall the sentences learned. We choose for two minutes of recall time because this is the amount of time used in a similar research by Godden and Baddeley in 1975 [KF2]. They gave the participants two minutes time to recall the learned words in the given context, which was underwater and on land. The participants must say the sentences out loud while their voice is being recorded by the supervisor using a mobile phone. Participants are prohibited to ask the supervisor for hints regarding the memorized sentences.

3.4.3 Assessment

The performance will be measured in terms of the number of sequence and omission errors. It is possible that the participant recalls a sentence differently than learned, but the meaning of the sentence is nearly the same. For this reason we use an online tool whereby the meaning of the sentence will be taken into account during the similarity measurement. The online tool that we use is provided by explosion.ai [19] and gives scores based on the percentage of similarity. For example the sentence is “I walk to the station” and the participant recalls the following sentence “I **walked** to the station”, then the score of that sentence will be 0.97 according to the similarity assessment tool. For the word list and anagram we assess by the amount of rightly given answers. For example the word list is “Flatness, Iron, Harbor, Crab, Thief” and the participant recalls the word list as follows “Flatness, **Cat**, Harbor, Crab, Thief” then the participant recalled four out of five words correct which results in the score of 0.8. During the test it is possible that the participant gives a word that is close to the learned word, then we asses this differently. For example, if the participant says the word list as follows “**Flat**, Iron, Harbor, Crab, Thief” whereby instead of ‘flatness’ ‘flat’ is being submitted, we asses this as a half correct word. This sequence will result in the score of 0.9. It is also possible that the participant might say the intended sentence with slight variance every time. We decided to only assess the last verbal submission to prevent lucky guesses. We give every sentence a score between zero and one. If a participant can not remember a sentence, then the sentence is assessed with a 0.0. The final score for the rote-learning task is the average of the scores of all the sentences.

3.5 Association Task

3.5.1 Memorizing

After the rote-learning task, the participants are asked to do an association test where they have to say out loud the first word that comes to mind in reaction to a given low association word, which is a word with no direct association. For example, the low association word is ‘happiness’; the participant says, “The given word is *happiness*” and then states “the word that comes to mind is *holiday*”. After the participant has said the associated word out loud, the participant has to go directly to the next given word. The ten low association words were fetched from the research of Burke *et al.* in 1987, and meet the word association requirements. In the task, the original words from Burke *et al.* are used for the English participants (see Figure 3.8) and directly translated to Dutch for the Dutch speaking participants (see Figure 3.9).

- Chance	- Money
- Cruel	- Now
- Lazy	- Size
- Melt	- Time
- Narrow	- Tall

Figure 3.8: Ten English low association words used in the association task

- Kans	- Geld
- Wreed	- Nu
- Lui	- Maat
- Smelten	- Tijd
- Smal	- Lang

Figure 3.9: Ten Dutch low association words used in the association task

3.5.2 Recall

On the day of recall, the participant has to do the same procedure as twenty four hours earlier. The participant will see the same stimulus words in the given context; they must recall the self-generated words and say them out loud just like the day before.

3.5.3 Assessment

The association task will be assessed in terms of right answers given by the test person. In this task, only the final submission for each given word will be assessed and the previous submissions will not be taken into account. The score will consist of the amount of correct answers. For example, if a test person recalled seven out of ten words correctly, then the score will be 0.7. Just like in the rote-learning task, we will assess words that are similar with half the points. For example, the self-generated word was 'happiness' and the next day the submitted word is 'happy', then we assess this word as half correct. The score of the task will always be between zero and one.

3.6 Recognition Task

The last test that the subjects are asked to complete is the recognition test. In this test, the test persons are asked to remember twenty different pictures.

3.6.1 Memorizing

On the day of memorizing, the test person is asked to memorize twenty different pictures. One picture per page will be shown in the virtual book or in the real book. The pictures are divided into two groups: one group with ten emotional pictures and the other group with ten neutral pictures. We based the selection of the neutral and emotional pictures on the research of Goodwin *et al.* from 1969 [KF1]. The pictures that are to be considered as neutral are mail order catalog models while the emotional pictures are cover models from erotically tinted magazines. The deviation of the male and female models in the categories emotional and neutral is 50:50. The sequence of showing the pictures to the test person is random. The test persons are allowed to scroll back with a time limit of five minutes.

3.6.2 Recall

The next day, the test persons are asked to select the twenty memorized pictures out of the now forty shown in the book. The test persons are allowed to scroll through the images and finally recite a maximum of twenty different page numbers which they think are the same twenty pictures seen twenty four hours prior. The maximum amount of time given for the submission is ten minutes. They are also allowed to say less than twenty numbers if the test person does not recognise the pictures. The twenty newly added pictures are similar pictures to the original twenty and have the same deviation in the categories- namely neutral and emotional. The pictures will be, just like twenty fours hours prior, each displayed on a whole page.

3.6.3 Assessment

The recognition task is assessed on the amount of correctly chosen pictures by the test persons. If the test persons change their mind, then the last submission will be seen as the final submission. If the test person says less than twenty numbers out loud, then they will not be punished on that, but it will reflect in the final scoring. For example if a participant names the ten rightly numbers and leaves the other ten, then the score will be ten out of twenty, so 0.5. If a participant names twenty numbers, of which ten are correct, then the score will be the same as named before, namely ten out of twenty, thus 0.5. The score of the recognition task will always be between zero and one.

3.7 Pilot

Before we started the experiment, we had four test persons executing the full test to see if the test needed adjustments. Three of the test persons are native Dutch and one of them is native English. During the pilot, we came to new insights. We saw that some Dutch participants were having difficulties performing the experiment in English. We thought that the lack of fluency in a language could affect the results of the experiment, therefore we made two versions of the test, namely in English and in Dutch. We also looked at aspects like the clarity and robustness of the virtual reality setup. We changed the explanatory text in such a way that it couldn't be misunderstood. During the test with the virtual reality setup, we also made a few modifications. For example, we made the font of the text in the virtual book larger. We also removed the physical stand with the book on it because the testers kept hitting it with the HTC Vive controllers. By changing the font-size of the text in virtual reality, we decided to make the font-size of the text in the reality setup the same size to keep consistency in our research. Another change to the experiment was regarding the demographic questions that the test persons had to answer on a form. At first we let the test persons answer the questions themselves on a piece of paper, but we came to the conclusion that this took too much time. We aimed at a maximum test time of 10 to 15 minutes total, so we decided that we would ask the questions verbally and submit the answers in Excel.

4. Results

In this section we are going to discuss the most relevant results according to this research. For the full results, see appendix A. We gathered a total of 57 test persons in 7 weeks, of which 6 did not show up to recall the next day. This leads to 51 full results to be examined. All the test persons tested are Dutch, which is good for the homogeneity of the test group.

During the test, we had multiple factors that could influence our gathered data. We took the results of the rote learning score, the association score, and the recognition score to make an overall score, which is the average of the previously named three scores. First of all, we see in Figure 4.1 that there is no significant difference between the overall scores of the females and the males who took the test. The p-value from the independent Student's T-test is 0.516 and thus not statistically significant. We can see that the males scored a little bit higher with an average overall score of 0.660 in relation to the score of 0.635 by the females. The experiment took place at two different locations, namely in the attic of a Dutch outdoor sporting goods store called *Bever*, and at a designated testing location at the Rotterdam University of Applied Sciences (HRO). To see if there is a significant difference in the overall scores of the participants between the two locations, we performed an independent Student's T-test. As the p-value of 0.641 shows in Figure 4.1, there is not a significant difference between the two groups, which means that there was no influence of the location of the setup on the overall scores of the participants.

Before the start of the experiment, we asked all the testers the title of their current occupation and if they already had experience in virtual reality. We want to see if there is a deviation of scores between students and waged staff. The results of an independent Student's T-test in Figure 4.1 shows that there is no statistical difference between the two groups and thus, the ten test persons of which their current occupation is waged staff, do not influence the results. If we look at if previous experience in virtual reality had influence on the results we can see in Figure 4.1 that this does not influence the overall scores (p = 0.839, independent samples T-test). This means that there are no biased test persons which could have had an advantage during the test.

Group	*n*	Mean overall score	*p*
Female	21	0.635	0.516
Male	30	0.660	
Bever	19	0.638	0.641
HRO	32	0.657	
Students	41	0.646	0.722
Waged staff	10	0.664	
Previous experience in VR	27	0.647	0.839
No previous experience in VR	24	0.653	

Figure 4.1: Table showing the results of multiple independent Student's T-tests

We also looked into the fact if the time between memorizing and recall affects the overall score of the test persons. We had in total 51 test persons who participated in our research and not all of the test persons performed the recall test exactly twenty four hours after the memorizing phase. We can see the correlation plot in Figure 4.2 that shows us that there is no correlation between retention time in minutes and overall score.

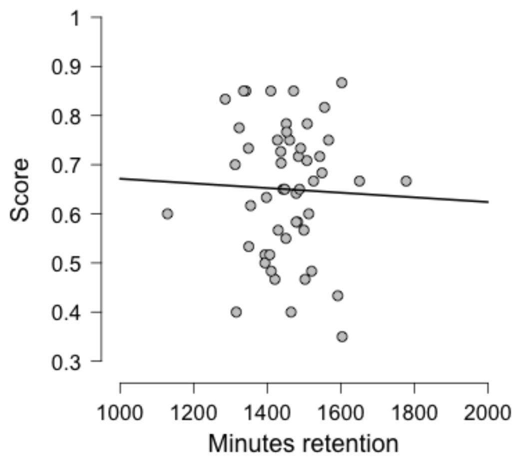


*Figure 4.2: Correlation plot showing the correlation between time retention in minutes and overall score (*n = *51,* r = *-0.037,* p *= 0.797)*

During the experiment we assigned each participant randomly to a designated context group. The context groups are the following: Reality - Reality (RR), Virtual Reality - Virtual Reality (VV), Virtual Reality - Reality (VR), and Reality - Virtual reality (RV). The first word in the hyphenated group title is the context in which the participant memorizes. The second word in the title represents the context in which the participant recalls.

In Figure 4.3 we can see the average overall scores per group plotted. According to the right figure in Figure 4.3 we can see that the group with no context change (combining groups RR and VV) scores significantly higher than the group with a context change (combining groups RV and VR). We can also see in the left of Figure 4.3 that the VR group scores the lowest of the four groups, which is quite striking.

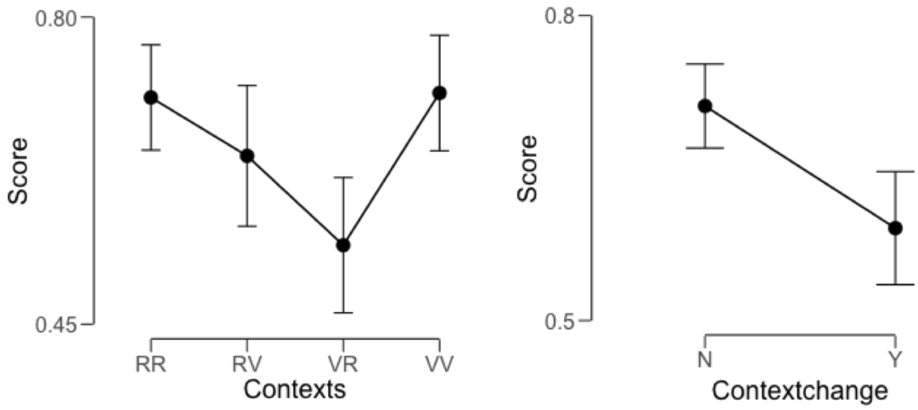


Figure 4.3: Plots showing the mean scores for different context categories

The overall score values within both context-change groups (RR - VV and RV - VR) can be considered to be distributed normally, according to a Shapiro-Wilk test (p = 0.278 for RR - VV, p = 0.794 for RV - VR).

Table 4.4 shows the results of multiple independent T-tests comparing different context groups. Most importantly, a change of context significantly affects the overall score in a negative way, when compared to subjects who experienced no change of context. This result suggests that a change in context has a negative influence on the memory.

<table>
<tr><th>Context group</th><th>n</th><th>Mean overall score</th><th>p</th></tr>
<tr><td>RR/VV (no context change)</td><td>25</td><td>0.711</td><td rowspan="2">< 0.001</td></tr>
<tr><td>RV/VR (context change)</td><td>26</td><td>0.591</td></tr>
<tr><td>RV</td><td>13</td><td>0.642</td><td rowspan="2">0.058</td></tr>
<tr><td>VR</td><td>13</td><td>0.540</td></tr>
<tr><td>RR</td><td>12</td><td>0.708</td><td rowspan="2">0.901</td></tr>
<tr><td>VV</td><td>13</td><td>0.713</td></tr>
</table>

Table 4.4: Table showing the results of multiple independent Student's T-tests, comparing different context groups

You can see in Figure 4.4 that the mean overall score of the group VR and RV are quite far apart from each other when you compare them. When you look at the groups RR and VV, we see that they are actually close to each other. The Student's T-test showed us that the groups RV and VR are almost statistically significant to each other with a *p*-value of 0.058, which concludes in a non-statistical effect. If we look at the other two groups, VV and RR, we see that the *p*-value is not statistically significant at all.

Task	***n***	**Mean overall score**	***p***
Rote-learning (*no context change*)	25	0.473	0.006
Rote-learning (*context change*)	26	0.294	
Association (*no context change*)	25	0.792	0.106
Association (*context change*)	26	0.715	
Recognition (*no context change*)	25	0.868	0.032
Recognition (*context change*)	26	0.763	

Table 4.5: Table showing the results of multiple independent Student's T-tests, comparing context change to no context change in a task

When we look if context change has an effect on the scores per task, in Table 4.5, we see that the overall scores of the rote-learning and the recognition task are significantly affected by the context change. On the contrary, there is not a significant effect on the association task.

5. Conclusion

From the gathered results, we saw that the scores from the test group were not influenced from factors like current occupation, sex, time retention, experience in virtual reality and location of the setup. All the participants were native Dutch speakers, which means that there could not be an influence of language on the scores. We can conclude that the gathered results are valid to draw conclusions from.
Our hypothesis, stated in the beginning of our paper, is that the change in context between virtual reality and reality has an effect on the ability to recall. In the previously discussed results, we saw that there is a highly significant difference in scores between the groups with context change and the groups without context change ($p < 0.001$). The group with a context change scored significantly lower than the group without context change.
The outcome of these results imply that it is more debatable if we can use learning in virtual reality in real life situations whereby the ability to recall plays a role. We gave an example in the beginning of this work whereby a military rescuer learned the positions of the hostages in a virtual environment to remember where the hostages are in real life. Does the found effect in this paper mean that it is impossible to use virtual reality in situations like the example? We can not answer this question with certainty. However, the discovered effect does mean we are less certain that virtual reality can be used for situations whereby recall in reality is important.

When we look back at Table 4.5, we see that the overall scores of the rote-learning task and the recognition task are significantly affected by the change in context, while there is no significant difference in the overall scores of the association task. In the association task, the participant creates self-generated content which needs to be memorized. We think that the overall score from the association task is less influenced by the context change because the association task is more focused on generating the content than on learning the content like in the other two tasks, but we can not say this with certainty. It is possible that self-generated memory is less sensitive to the change in context effect. We can conclude that not all forms of memory are equally sensitive to the effect.

We also found a result during this experiment which remains unexplained. We saw that the VR group scored the lowest of all groups on the test. When we compared these scores to the RV group, it resulted into a *p*-value of 0.058, which is close to a statistical significance. Our hypothesis is that learning in the virtual reality context is more difficult than in the reality context. We think that elements like excitement for virtual reality and the unknowingness of the contents of the test makes it more difficult for the participant to learn in the virtual reality context. We think that the recall in virtual reality is easier than memorizing because the participant already has an idea what is going to happen in the test and thus the unknowingness of the test is lessened, while the excitement for the virtual reality is still there. If our hypothesis is correct about this non-statistical effect, then we should also see that the VV group scores lower than the RR group. When we look back at Figure 4.4, we see that the VV scores even a bit higher than the RR group, which makes our hypothesis invalid. We can only tell that there is an non-statistical effect between the RV and VR group.

5.1 Future Recommendations

As discussed above, we found an almost statistical difference between the RV and VR group whereby VR scored the lowest. The effect stays unexplained and needs further research into the matter even though the result is statistically insignificant.
Another matter that is worthy to look into is that, in this paper, we only looked at the short-term consequences of context-dependent memory in relation to virtual reality and did not take long-term memory into account. Future research is advised to be done to see what the long-term effect is of context-dependent memory in virtual reality. Another matter that needs further investigation is if the graphic and immersive quality of the virtual reality affects the context-dependent memory. During this research we did not have enough time and man-power to investigate this matter because the amount of development time and participants would increase exponentially. We could hypothesise that if the virtual world looks as close as possible to the real world in terms of graphics and immersiveness that the discovered effect in this paper, whereby the change in context negatively influences the ability to recall, is less dominant.

6. Acknowledgements

I thank Bever Leiden, and especially Astrid Overduin, gratefully for giving me the opportunity to install my test setup in their attic to use as a testing location. I would also like to thank my second advisor, Jasper Schelling, who helped me find a location within the Rotterdam University of Applied Sciences to perform my research. The Rotterdam University of Applied Sciences, and in particular the Communication and Multimedia Design program, deserves my special thanks for their warm hospitality and help. Whilst doing this research, I received profound support from my parents, sister, and girlfriend. Their endless compassion and encouragement enabled me to succeed in this venture. To them, I give my ultimate thanks.

References

1. Goodwin, D. W., Powell, B., Bremer, D., Hoine, H., & Stern, J. (1969). Alcohol and recall: State-dependent effects in man. Science, 163(3873), 1358-1360.
2. Godden, D. R., & Baddeley, A. D. (1975). Context- dependent memory in two natural environments: On land and underwater. British Journal of psychology, 66(3), 325-331.
3. Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. Memory & Cognition, 6(4), 342-353.
4. Bower, G. H., Monteiro, K. P., & Gilligan, S. G. (1978). Emotional mood as a context for learning and recall. Journal of Verbal Learning and Verbal Behavior, 17(5), 573-585.
5. Johnson, A. J., & Miles, C. (2008). Chewing gum and context- dependent memory: The independent roles of chewing gum and mint flavour. British Journal of Psychology, 99(2), 293-306.
6. Attree, E. A., Brooks, B. M., Rose, F. D., Andrews, T. K., Leadbetter, A. G., & Clifford, B. R. (1996, July). Memory processes and virtual environments: I can't remember what was there, but I can remember how I got there. Implications for people with disabilities. In ECDVRAT: 1st European Conference on Disability, Virtual Reality and Associated Technologies. Reading, UK (Vol. 118).
7. Dinh, H. Q., Walker, N., Hodges, L. F., Song, C., & Kobayashi, A. (1999, March). Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In Virtual Reality, 1999. Proceedings., IEEE (pp. 222-228). IEEE.
8. Plancher, G., Gyselinck, V., Nicolas, S., & Piolino, P. (2010). Age effect on components of episodic memory and feature binding: A virtual reality study. Neuropsychology, 24(3), 379.
9. Gould, N. F., Holmes, M. K., Fantie, B. D., Luckenbaugh, D. A., Pine, D. S., Gould, T. D., ... & Zarate Jr,, MD, C. A. (2007). Performance on a virtual reality spatial memory navigation task in depressed patients. American Journal of Psychiatry, 164(3), 516-519.
10. Sweeney, S., Kersel, D., Morris, R. G., Manly, T., & Evans, J. J. (2010). The sensitivity of a virtual reality task to planning and prospective memory impairments: Group differences and the efficacy of periodic alerts on performance. Neuropsychological Rehabilitation, 20(2), 239-263.
11. L. E. Marks and G. A. Miller, J. Verb. Learn. Verb. Behav. 3, 1 (1964).
12. Burke, D. M., Peters, L., & Harrold, R. M. (1987). Word association norms for young and older adults. Social and Behavioral Science Documents, 17(2).
13. Chomsky, N. (1956). Three models for the description of language. IRE Transactions on information theory, 2(3), 113-124.
14. Descriptors defining levels in the European Qualifications Framework (EQF) - Learning Opportunities and Qualifications in Europe - European Commission. (n.d.). Retrieved March 19, 2018, from https://ec.europa.eu/ploteus/en/content/descriptors-page
15. Standardized Survey Classifications - Individuals. (n.d.). Retrieved March 19, 2018, from http://www.pgagroup.com/standardized-survey-classifications.html
16. Bitmill Inc. (n.d.). Word Lists. Retrieved March 19, 2018, from https://www.litscape.com/word_tools/anagram.php

17. Random Word Generator - Creative online tool to generating randomized words for brainstorming. (n.d.). Retrieved March 19, 2018, from https://www.textfixer.com/tools/random-words.php
18. Anagrammen. (n.d.). Retrieved March 19, 2018, from http://www.mijnwoordenboek.nl/anagrammen/
19. Sentence Similarity. (n.d.). Retrieved March 19, 2018, from https://demos.explosion.ai/similarity/?text1=This is a sentence&text2=This is another sentence

Appendix A

Independent Samples T-Test Location

Independent Samples T-Test

	t	df	p
Score	-0.470	49.000	0.641

Note. Student's t-test

Descriptives

Group Descriptives

	Group	N	Mean	SD	SE
Score	Bever Leiden	19	0.638	0.157	0.036
	HRO	32	0.657	0.121	0.021

Independent Samples T-Test Sex

Independent Samples T-Test

	t	df	p
Score	-0.655	49.000	0.516

Note. Student's t-test

Descriptives

Group Descriptives

	Group	N	Mean	SD	SE
Score	Female	21	0.635	0.146	0.032
	Male	30	0.660	0.126	0.023

Independent Samples T-Test Current Occupation

Independent Samples T-Test

	t	df	p
Score	-0.358	49.000	0.722

Note. Student's t-test

Independent Samples T-Test Experience VR

Independent Samples T-Test

	t	df	p
Score	0.136	49.000	0.893

Note. Student's t-test

Independent Samples T-Test Context Change

Independent Samples T-Test

	t	df	p
Score	3.551	49.000	< .001

Note. Student's t-test

Assumption Checks

Test of Normality (Shapiro-Wilk)

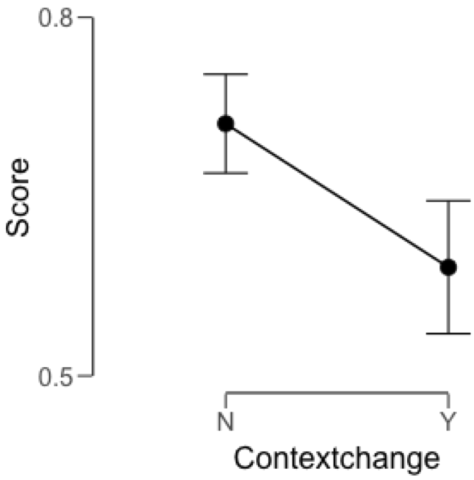
		W
Score	N	0.952
	Y	0.977

Descriptives

Group Descriptives

	Group	N	Mean	SD	SE
Score	N	25	0.711	0.100	0.020
	Y	26	0.591	0.138	0.027

Descriptives Plot
Score



Independent Samples T-Test
RV-VR

Independent Samples T-Test

	t	df	p
Score	1.994	24.000	0.058

Note. Student's t-test

Assumption Checks

Test of Normality (Shapiro-Wilk)

		W
Score	RV	0.975
	VR	0.968

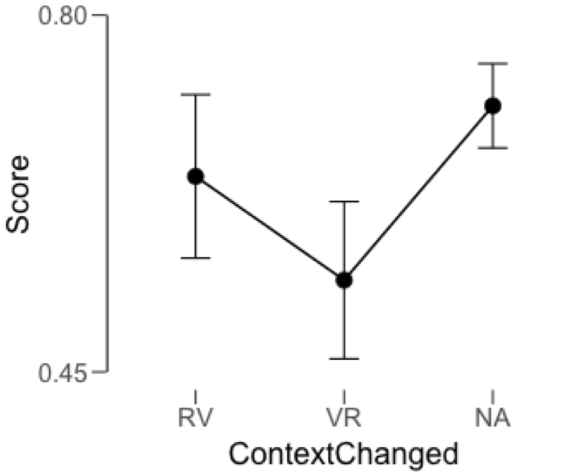
Note. Significant results suggest a deviation from n

Descriptives

Group Descriptives

	Group	N	Mean	SD	SE
Score	RV	13	0.642	0.133	0.037
	VR	13	0.540	0.128	0.035

Descriptives Plot
Score



Independent Samples T-Test
RR-VV

Independent Samples T-Test

	t	df	p
Score	-0.125	23.000	0.901

Note. Student's t-test

Assumption Checks

Test of Normality (Shapiro-Wilk)

		W
Score	RR	0.921
	VV	0.922

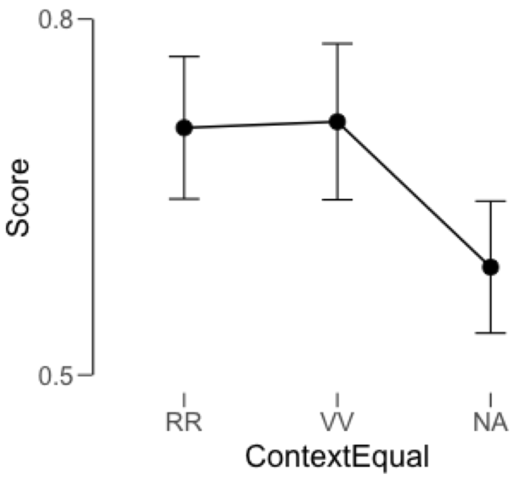
Note. Significant results suggest a deviation from n

Descriptives

Group Descriptives

	Group	N	Mean	SD	SE
Score	RR	12	0.708	0.094	0.027
	VV	13	0.713	0.109	0.030

Descriptives Plot
Score



ANOVA

ANOVA - Score

Cases	Sum of Squares	df	Mean Squar
Contexts	0.251	3.000	0.084
Residual	0.647	47.000	0.014

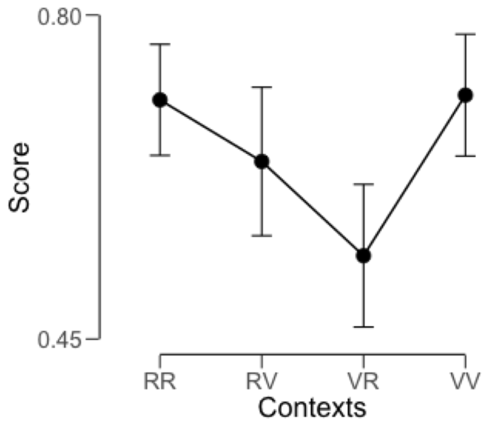
Note. Type III Sum of Squares

Descriptives

Descriptives - Score

Contexts	Mean	SD	N
RR	0.708	0.094	12.000
RV	0.642	0.133	13.000
VR	0.540	0.128	13.000
VV	0.713	0.109	13.000

Descriptives Plot



Descriptives

Group Descriptives

	Group	N	Mean	SD	SE
Score	N	24	0.653	0.142	0.029
	Y	27	0.647	0.130	0.025

Descriptives

Group Descriptives

	Group	N	Mean	SD	SE
Score	Student	41	0.646	0.128	0.020
	Waged staff	10	0.664	0.163	0.052

Independent Samples T-Test - Context change

Independent Samples T-Test

	t	df	p
Rote-learning score	2.863	49.000	0.006
Association score	1.646	49.000	0.106
Recognition score	2.206	49.000	0.032 [a]

Note. Student's t-test

[a] Levene's test is significant ($p < .05$), suggesting a violation of the equal variance assumption

Descriptives

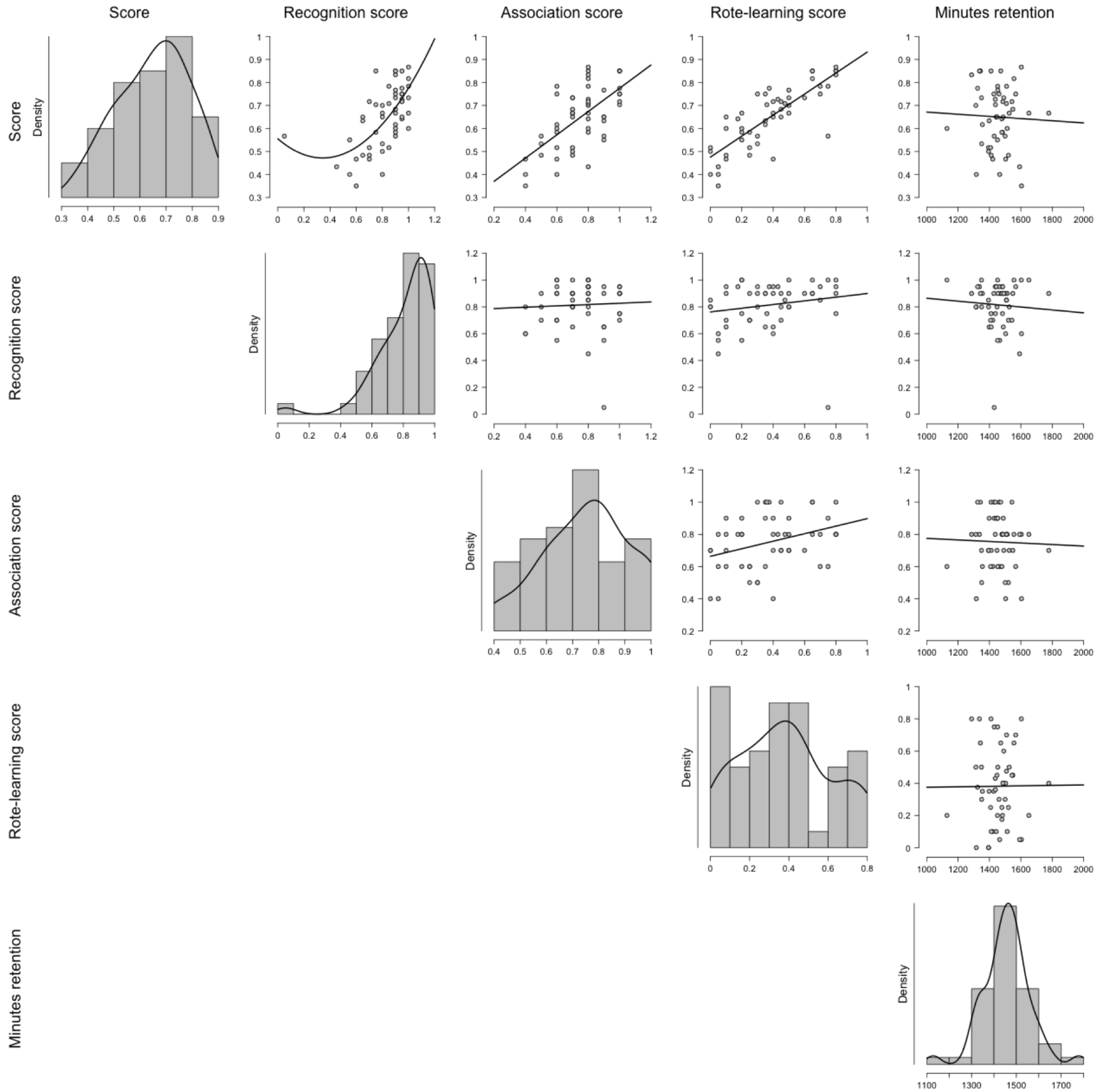
Group Descriptives

	Group	N	Mean	SD	SE
Rote-learning score	N	25	0.473	0.221	0.044
	Y	26	0.294	0.225	0.044
Association score	N	25	0.792	0.144	0.029
	Y	26	0.715	0.185	0.036
Recognition score	N	25	0.868	0.104	0.021
	Y	26	0.763	0.214	0.042

General Descriptives Statistics

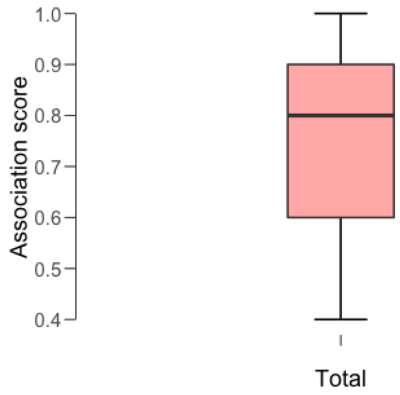
	Score	Recognition score	Association score	Rote-learning score	Minutes retention
Valid	51	51	51	51	51
Missing	0	0	0	0	0
Mean	0.650	0.815	0.753	0.382	1455.706
Median	0.667	0.900	0.800	0.375	1453.000
Std. Deviation	0.134	0.176	0.169	0.239	104.341
Minimum	0.350	0.050	0.400	0.000	1129.000
Maximum	0.867	1.000	1.000	0.800	1778.000

Correlation plots

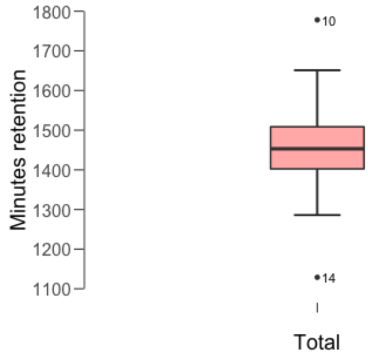


Boxplots

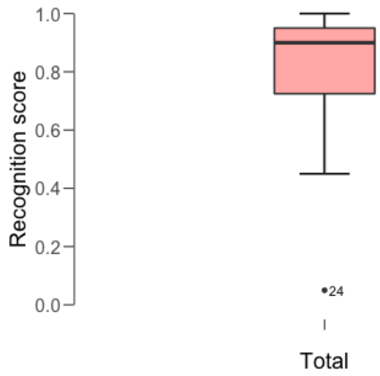
Association score



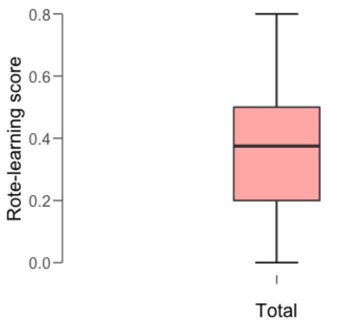
Minutes retention



Recognition score



Rote-learning score



Score