Tutorial Adaptation based on Working Memory[1]

Miguel Maria do Nascimento Simões Mortágua Keim

Thesis to obtain the Master of Science Degree in

Computer Science and Engineering

Supervisors: Engª Marta Barley de La Cueva Couto
Prof. Carlos António Roque Martinho

Examination Committee

Chairperson: Prof. Daniel Simões Lopes
Supervisor: Engª Marta Barley de La Cueva Couto
Member of the Committee: Profª. Joana Carvalho Filipe de Campos

October 2023

Acknowledgments

To start off, I'd like to thank my two supervisors- Prof. Carlos Martinho and Profª. Marta Couto- for guiding me through the whole process of this project. It is thanks to their expertise that it has come this far, with Prof. Carlos providing me with the necessary knowledge in the field of Game Design and Programming, and Prof. Marta supporting me with all the information regarding psychology and associated studies needed for my research.

Of course, I would like to give a special thanks to my mother, who has tirelessly worked to sustain a comfortable lifestyle for both me and my sister, even through the loss of our father, and to whom I owe this chance in the first place. None of this would be possible without her ever present support, and no amount of words can express all that she has done, both in raising me and helping me throughout my life, studies or otherwise.

I extend my gratitude to all my wonderful friends as well, of course, for providing me a source of motivation, support and happiness needed to continue my journey. And last but not least, I want to express my gratitude to the LabJogos association in Instituto Superior Técnico Taguspark for extending their help when I was in need of participants for my project. Without them and my friends, I would not have gone this far in it, and I'm most grateful.

To each and every one of you – I thank you.

Abstract

With the popularity of video games growing exponentially over the years, the complexity and diversity of the genres and their games continues to expand. Certain problems begin to arise, however- with complex mechanics comes a need for good tutorials and proper ways to convey the information the player needs in the most optimal and concise way possible. We believe the key to solving this problem lies in its connection to our memory, the most integral part in the gathering of new knowledge.

To achieve this, we import concepts from cognitive psychology and test two approaches on participants with different Working Memory capacities. We will test two opposite scenarios, where we apply different types of learning experiences in a Tactical Role-playing Game: (1) we drew inspiration from the Generation Effect to design a tutorial based on exploration and limited information, where the player must seek answers and gather knowledge on their own and (2) a didactic tutorial, where the game teaches in detail the mechanics of the game to the player, and how to tackle the scenario in front of them.

By measuring levels of Working Memory through a test, and then dividing our participants into the different tutorial environments, we aim to evaluate their retention of information in their respective scenario, as well as their subsequent performance in the following levels of the game. This way, we can draw potential conclusions on which tutorial approach is best adapted to take different levels of learning into consideration, or if certain methodologies work better for a specific type of player. While we were unable to gather enough of a sample to make a proper conclusion, we laid the groundwork for future implementations of the idea.

Keywords

Working memory; Long-term memory; Generation effect; Game development; NASA TLX.

Contents

List of Figures

Acronyms

CL Cognitive Load

CST Complex Span Task

GC Generation Constraint

GE Generation Effect

LTM Long-Term Memory

MM Memory Model

STM Short-Term Memory

WM Working Memory

1

Introduction

Contents

In this chapter, we will explore the incentive behind this work, the problems that we wish to resolve and the hypothesis behind our solution- in this case, we wish to explore the role memory plays in the learning of the evermore complex video game mechanics of today's day and age, and how to best utilize that knowledge to fabricate the most optimal tutorial for the experience, regardless of individual differences. At the end of this section an outline for the document will be provided for further clarification of the chapters that constitute it.

1.1 Motivation

As the years go by, technology progresses, and with it, so do video games. Gone are the days when simple games like "Pon"[1] and "Tetris"[2] ruled the market, and with each passing generation of both hardware and software, more complex options are available for game designers. Said evolution in technology brings forth more advanced experiences with more in-depth mechanics- and with these ever-growing changes, players must adapt and learn to enjoy themselves properly.

How often do we come across games with complicated menus, with more options than we can account for? This is especially true in the case of slower strategy games, where the player is encouraged to stop, think and check as many details as possible. However, it also applies to more active games as well, such as the Souls Series games like "Dark Souls"[3] and "Sekiro"[4], where the player must learn to master a variety of controls in fast paced action, under highly stressful situations. Such extensive arsenals can be overwhelming, and the responsibility of properly conveying such information lies on the shoulders of tutorials, that many times go unappreciated within the community.

It is integral to any good game to have an appropriate tutorial that feels fluid with the gameplay to avoid dragging down the experience for the players. First impressions are always important, and the faster a player can get a hold of how the mechanics of a game works, the better. While research has been done to try and get to the bottom of it, with studies breaking down the proper structure of tutorials [1] to figure out the best approach for video games, not many studies focus on the deeper mechanisms of knowledge retention in the field of video games. The ability one has to retain the information provided to them is linked to their memory capacity, and tutorials are not adapted to take that into consideration.

[1] Pong, 1972, Atari, Arcade
[2] Tetris, 1988, Mirrorsoft, Electronika 60
[3] Dark Souls, 2011, From Software, PlayStation 3
[4] Sekiro, 2019, From Software, PlayStation 4/Windows/Xbox One

1.2 Problem

Tutorials are a crucial part of most games, responsible for letting players dip their toes in the complex mechanics of the ever-evolving games of the present market. Game developers often disregard the importance of a good tutorial, instead overwhelming their players with large quantities of information, while not letting them absorb previously given pieces of the puzzle that is their game. Tutorials should be concise and direct to the point, and know how to space new mechanics to allow gamers to practice their newly acquired knowledge, until it is properly retained within their memory.

But not everyone learns at the same pace- that much has become clear with more recent findings in the field of psychology and memory. We each have our own capacity for retaining new information and learning speeds, so how exactly are we supposed to take that into account when planning how to teach our target audience? To our knowledge, no studies have researched it thus far, at least not specifically in game design. It should be possible to have adapted tutorials capable of taking different learning capabilities into consideration, or even find a specific type of tutorial best suited for any type of player we might come across.

1.3 Hypothesis

For us to be able to adapt tutorials, regardless of the individual learning differences we have as people, we must understand how the Human Memory works, a system we all share regardless of such variances. The key lies in our **Working Memory (WM)** [2]- the cognitive system responsible for retaining new information temporarily- which holds the answers to the aforementioned disparity in learning capabilities, due to its varying capacity. Said capacity, besides changing from individual to individual, is also limited, and will fade if not properly trained, like any muscle in our body. In order for the information it stores to be converted into **Long-Term Memory (LTM)**, in other words, for it to become permanent knowledge, it must be repeated and exercised, because the more a certain information is accessed, the stronger the neural network related to it gets.

It stands to reason that tutorials are our starting point as players in video games, and not just because they are the literal beginning of the experience, but because they are the introduction to the game's core mechanics. As such, they are responsible for optimizing the absorption of information to the LTM storage, allowing players to retain new knowledge at an appropriate pace, before jumping into situations that will test what players have learned. However, our capacity for knowledge retention is directly linked to our WM capacity, since it is what allows us to process new information and temporarily contain the data presented for it to be trained. That is what we aim to study through this research- how to adapt tutorials to take into consideration different WM capacities, in order to lessen the impact it holds on a player's experience.

The first step we must establish is how to estimate the individual differences in WM in the first place. Experiments regarding measures of WM have already been developed in the field of psychology [6], and thus can be applied here as well. Suppose we can use them to quantify our WM capacity at the beginning of a video game experience. In that case, it is possible to utilize this information to filter our participants at both ends of the spectrum to different types of tutorials. By dividing both those with high WM capacity and the ones with low WM capacity equally through the scenarios, we can verify the effects WM has in retaining information, as well as test different learning strategies in digital games, for us to establish which provides better retention of the information presented in either type of players.

In order to choose our tutorial environments for this test, we must explore different approaches to the same goal: how do we maximize the retention of information in our players at the start of a digital game experience? Learning is a broad and very explored subject in the area of Education, with studies already reaching conclusions to the question posted above. In the field of education, research has found that self-generated knowledge provides better long term results [10], albeit the exact reason behind such an effect is still unclear to this day. One such approach that focuses on self-generated learning is called the **Generation Effect (GE)**.

This method tells us that by providing the tester with incomplete information, as opposed to giving them all of the details, we can enhance their retention of the knowledge provided, since they will be forced to complete said information and exercise the created memory. This effect has been proven effective, albeit slower in the beginning when compared to more didactic approaches, especially in complex scenarios. Such theories have, to our knowledge, never been tested in the field of digital games, and we believe it holds unexplored potential to further enhance learning experiences in video games, especially since research has shown its connection to the use of mental resources [11]. We wish to apply it in a video game of our own creation, and test its relationship with WM when compared to a didactic tutorial, where all the information is provided to the player, akin to a lecture in a classroom.

After the tutorials, participants will be tasked to play additional levels of the game we will provide, where no more help will be given and they will have the opportunity to utilize the knowledge previously obtained in the learning stage. This is when we will be able to see the performance of the different levels of WM, when applied to the different learning approaches, in a real scenario. To finalize the process, we will also measure our participants' **Cognitive Load (CL)** through the NASA TLX [16] survey, since CL refers to our usage of WM resources to perform tasks, and will help us further understand the impact both tutorial approaches have in individuals with different WM capacities [19].

We hypothesize that, by utilizing the information gathered through this experiment, we will be able to understand how we could adapt tutorials to fit our players while taking into account their WM capacity. We believe the WM capacity should influence how quickly our players will familiarize themselves with the mechanics presented to them, while the GE affects how well one retains the new knowledge they have acquired [10], even if initially slower in results when compared to a more lecture-like tutorial, depending on the mental effort required [11]. We also expect players with high WM capacity to better adapt themselves to the GE tutorial, considering they should show more capacity for retention of new information [5], while the low WM ones should instead show better results in an environment best suited for their lower capacity for retention, like the didactic approach, which provides more guidance. Regardless, we aim to verify how both types of players behave in both types of environments in order to draw more concrete conclusions.

1.4 Document Outline

This document will be structured as follows:

- **Chapter 1: Introduction**, where we provide our readers with the main motivation for this work, the problem we seek to solve and the solution we hypothesize. In this case, how we can leverage knowledge about Working Memory and the Generation Effect to adapt tutorials to different types of players.

- **Chapter 2: Background**, a section dedicated to giving our readers the necessary knowledge to properly understand key concepts explored in this work. This includes the studies of Human Memory, such as Working Memory, Long-Term Memory and Cognitive Load, including the most widely accepted Memory Model, as well as different types of learning procedures, including an unexplored topic in video games, the Generation Effect, and a theory regarding it which supports its connection to the Human Memory.

- **Chapter 3: Related Work**, highlights other studies done in a similar vein throughout the years, be it related to video game tutorials or the workings of our memory and brain. Through it, we can understand other approaches in the field and gather information that can support our work.

- **Chapter 4: Implementation**, in which we detail our approach on the problem, the video-game environment in which our experiment will be structured upon, from its inner workings to how it was defined to better suit this work's intended objectives, and finally how we chose to gather data and evaluate it in face of the subject at hand.

- **Chapter 5: Evaluation**, where we showcase the structure utilized for the experiment ran with our participants, the results derived from it and the hypothesis derived from analyzing such data.

- **Chapter 6: Conclusion**, the final section of this research, in which we go over everything discussed previously and summarize it all for the sake of clarity and exposition, as well as go over potential limitations and future prospects from this work.

2

Background

Contents

In this chapter, we focus on key concepts of our study. We will briefly explain the most accepted model of Human Memory [5] and detail the differences between Long-Term Memory and Working Memory. Following this, we will explain how we can measure Working Memory capacity. We will also detail what the Cognitive Load is, which will be used in the final stages of our work, and finally, we will describe the Generation Effect and the evidence supporting it, as well as entering into detail on a specific theory which supports a connection to other topics presented.

2.1 Long-Term Memory

LTM was first defined by Atkinson-Shiffrin [2] in his multi-store memory model, back in 1968. The LTM was responsible for the retention of information and skills for long periods of time- so long, in fact, that said memories are believed to potentially last a lifetime. It is said that our LTM holds unlimited capacity, and the main constraint people feel with the passage of time links instead to its accessibility. In order for information in the WM to become permanent, it must be rehearsed multiple times, each time strengthening its connection in the LTM, and likewise the longer it stays in short-term storage, the stronger its roots within our memory becomes. The transferring process between the two memory storages is called Synaptic Consolidation [3].

After training new information in our WM, it is integrated in our LTM within structures called Schemas- either stored in existing ones if associated with them, or by creating new ones. These Schemas are created by our brain in order to ease the burden of our WM: through the process of association, we can more easily understand new knowledge if we already have a Schema related to it. Through the same logic, novel information applies more of a burden in our WM, since we lack any previous information related to it. And after all this, it is still necessary to train our brain in order to prevent the forgetting process- even if our LTM is unlimited, it does not mean all the information will be kept without any maintenance. This is usually done through rehearsal, where we recall certain knowledge and practice it to clarify it within our memory.

2.2 Working Memory

Prior to the second half of the 20th century, there was no concept of separation between two different types of memory storage. It was only in 1957, when William Beecher Scoville and Brenda Milner published an article by the name of “Loss of recent Memory after Bilateral Hippocampal Lesions” [4], that it was finally considered. The duo found out through examination of patients with hippocampal lesions that, even if they were unable to add new information to their long-term memory, it was still possible for them to process immediate input on the short-term. This discovery was then further explored by Richard Atkinson and Richard Shiffrin [2], where they deepened the differences between the **Short-Term Memory (STM)** and LTM, which eventually created the first **Memory Model (MM)**- a representation of the inner workings of memory within our brain- known as the Multi-Store Memory Model. Other MMs have since been theorized, which to this date are still discussed among researchers. We will be touching upon the most pertinent MM to our research shortly.

It was Richard Atkinson and Richard Shiffrin's studies that gave birth to the concept of WM- also known as STM, albeit some scientists debate that the two should be separate. Our work will, however, focus on the WM as its own separate concept. This type of memory storage is a cognitive system that allows us to hold information temporarily. It is an integral part of our brain, responsible for handling a lot of our decision making. However, it is limited, as opposed to the aforementioned Long-Term Memory, which holds theoretically unlimited permanent knowledge capable of lasting for decades. The bigger one's WM capacity, the greater number of pieces of information they can hold and process at a time, and it is believed most adults can store between 5 and 9 items at a time within it. This information must then be rehearsed and practiced, as we have explained before, as to be converted into LTM and stored within one of the aforementioned Schemas, or if lacking any previous association to other stored memories, creating its own Schema.

2.2.1 Baddeley and Hitch's Memory Model

The most relevant MM to our project was proposed in 1974 by Baddeley and Hitch, called the "Working Memory Model" [5], responsible for further explaining the structure and process behind our STM. This model claimed that the WM was not a single structure but was divided into multiple subcategories, each responsible for different tasks (Fig. 2.1):

The **Central Executive** is the system responsible for the control and regulation of cognitive processes. Through it, we can direct focus and target specific information, as well as allocate the information received to its supposed system; The **Phonological Loop** is a temporary storage that deals with sound or phonological information. It is divided in two sectors: the Phonological Store, responsible for storing all the heard information, and the Articulatory Process, which rehearses any words or sounds to keep them in the WM; The **Visual-spatial Sketchpad**, which works with visual and spatial information, be it storing, processing or manipulating it; And finally, a final sector later introduced to the model in 2000- the **Episodic Buffer**. This component links information across domains to form integrated units of the other sectors in a time sequencing manner, much like memories of a story of a movie scene, akin to a buffer. It is believed that it also links LTM to its semantic meaning.

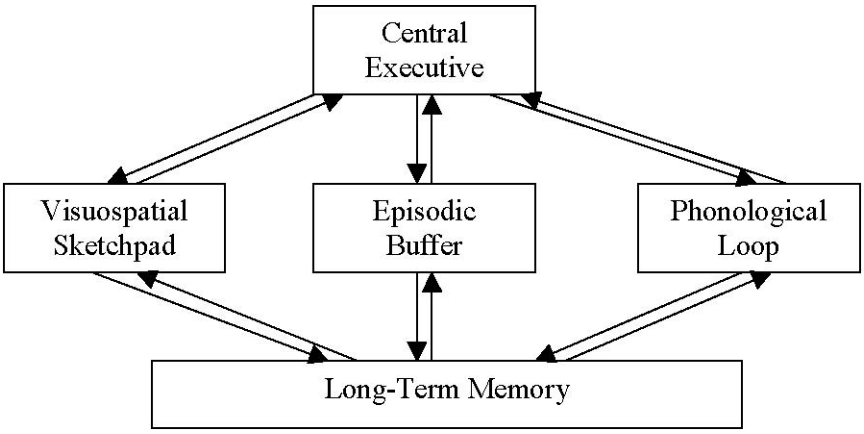

Figure 2.1: Baddeley and Hitch's Memory Model, 2000.

2.3 Working Memory Capacity Measures

The most widely utilized process to evaluate one's Working Memory Capacity comes in the form of **Complex Span Task (CST)**- exercises that mix memory tasks, such as remembering a sequence of objects, with interleaved secondary processing tasks, like judging the correctness of equations. These tasks are thus designed to engage multiple aspects of our working memory by targeting different brain regions, while limiting one's ability to utilize mnemonics through demanding processes. By utilizing the CST, researchers have been able to establish that the average of items lies around 4, when previously with Simple Span Tasks the number registered was 7 [8].

Many forms of CSTs have been explored throughout the years, with researchers aiming to further understand the mechanics behind such phenomena, as well as find out more accurate measures. The first example of a CST dates back to 1980, with the "Reading Span" by Daneman and Carpenter [6], where they combined the storing process of a Simple Span Task (for example, remembering a sequence of words) with a secondary task like reading the phrase presented. Other examples of relevant tests that were later developed based on this research would include the Operation Span by Turner and Engle in 1989 [7], where participants had to solve arithmetic equations while remembering a list of unrelated words, which proved that the processing content of the CST didn't need to be similar to measure WM; and the Symmetry Span in 1996 by Shah and Miyake [9], in which participants made symmetry judgments and remembered spatial locations, which demonstrated a verbal-spatial distinction when compared to the Reading Span in the prediction of spatial abilities.

2.4 Cognitive Load Theory

The **CL** corresponds to the amount of resources used from our WM at a time, as defined by John Sweller in 1980 [15]. The Cognitive Load Theory was developed with the intent of helping instructors optimize teaching designs to lower the toll on the CL of their learners, since those with lower WM capacity would struggle to keep up with new information when under heavy load. This theory, in essence, proposed that quality of instructional design would be increased if the limitations of WM were taken into consideration during the process.

Sweller divided CL into three different types: the Intrinsic Load, the Extraneous Load and the Germane Load. The **Intrinsic Load** refers to the load associated with the inherent difficulty and complexity of a given task, varying depending on the degree and number of complex concepts needed for it to be processed; the **Extraneous Load** is the load utilized for processing information that is dissociated from the learning of the current task, including the distractions that otherwise hinder it; and the **Germane Load**, the type of load which process the stored information in order to create the Schemas we have previously touched upon when talking about the LTM.

By combining these three varieties one would then be able to get the CL currently being used, an amount that shouldn't surpass the WM capacity of the individual in question, as that would impede the proper processing of newly acquired information. The most widely used measures of CL are subjective in nature, based on ratings of perceived mental effort and task difficulty, and not objective calculations. Some examples would include the NASA Task Load Index [16] and the 9-point Linkert scale utilized by Paas [17].

2.5 Generation Effect

Memory researchers have been investigating effective mnemonics- techniques developed to enhance the storage and retrieval of information from our memory- for several years, hoping to understand how such strategies can promote better memorization. Not only are they useful in our daily lives for remembering names and tasks, but they also hold a proven advantage in the educational field. These researchers hope that, by understanding the mechanics behind such strategies, we will further master our studies of memory as a whole.

One such strategy that has shown promise is known as the GE, first explored by Slamecka and Graft in 1978 [10], a phenomenon where information is better remembered if self-generated, as opposed to reading. This type of mnemonic has shown results through many experiments, often using word lists to further study different benefits of self-generated memories. These tests often presented the participants with a list of stimuli, usually pairs of words. Half of the participants would be given intact target words, while the other half would have to self-generate the pair with incomplete words instead[1]. Later stages of the experiment would find better memorization in the section of the group that had to self-generate their answers. That being said, even though many theories have been developed around this effect, there hasn't been a single hypothesis that has been able to satisfy all the questions regarding the mechanics behind the GE.

Furthermore, one factor within this effect has been explored in the hopes of justifying the strategy- known as the **Generation Constraint (GC)**- which refers to the amount of information an individual is given that limits what can be produced in such generated tasks. In essence, these constraints serve to filter the amount of possible answers a participant can give to the problem they face, working as clues to funnel them towards the expected answers. However, there have been works that have disproved the utility of such tactics, showing that testers with fewer constraints can generate better memory benefits. Regardless, studies keep testing various constraint levels to verify the extent to which the GC can influence both item and context memory.

2.5.1 Mental Effort Theory

Among the many theories revolving around the GE, there exists one most relevant to our studies. In 1979, Tyler et. al. [11] found a potential relationship between the GE and Mental Effort. Their theory suggested that self-generated information requires more mental effort than other methodologies, i.e., a larger amount of cognitive resources must be utilized to perform a given task. Their results showed that, for instance, high-effort self-generated information led to better retention than low-effort scenarios. Given that mental effort is, by definition, associated with our cognitive resources, this would imply a direct connection to our WM, and as such, this theory could serve as a backbone to prove a logical connection between the GE and our WM capacity which we aim to research in our paper. However, a lack of universal measures and manipulation of mental effort has held this theory back in use, since there is no reliable way to quantify the data it wishes to evaluate, and it has so far relied on subjective measurements instead.

2.6 Summary

In 1968, the LTM was first defined by Atkinson-Shiffrin in his multi-stored memory model. In his experiments, he discovered that a section of our memory was responsible for storing long-term information and skills, with a potentially unlimited capacity and for a long time. The more this information is trained, the longer it stays within our memory, in segments called Schemas- organized through association of

[1] e.g. Intact pair of keywords: Above-Below; Incomplete pairs: Open-C...

different key information. The more Schemas we have associated with a new piece of information, the easier it is for us to understand and retain that new knowledge.

On the other hand, the STM, or WM, was originally not separated from its long duration counterpart. It was William Beecher Scoville and Brenda Milner in 1957 that published the first article that considered the possibility, when they investigated patients with hippocampal lesions. This information was further explored by R. C. Atkinson and R. M. Shiffrin in their own research, creating the first MM, a representation of the inner workings of our memory. The newfound STM, unlike the LTM, is capable of holding information temporarily with a limited capacity, responsible for handling a lot of our decision making, and the first step towards the retention of new knowledge. The bigger our WM, the bigger the amount of information we can process at a time.

While not final, the most widely accepted MM was established in 1974 by Baddeley and Hitch, which divided the STM into multiple subcategories, each with their own designated responsibility. It contains a Central Executive, which regulates the remaining sections; the Phonological Loop, responsible for sound-related memories; the Visual-spatial Sketchpad, which as the name implies deals with visual and spatial information; and the Episodic Buffer, a subcategory later introduced as the one responsible for linking the other sectors.

Back in 1980, the most commonly used measurement of our WM was established by Daneman and Carpenter, through tests called the CSTs- exercises that mix memory tasks, such as remembering a sequence of objects, with interleaved secondary processing tasks, like judging the correctness of equations. Further experiments were developed on the subject, but they all revolved around this core concept, and have been proven to correctly work as a measuring tool.

John Sweller would define in 1980 the Cognitive Load Theory, with the goal of helping instructors optimize teaching designs in order to lower the toll on the CL of their learners- the amount of resources used from our WM at a time. Taking different WM levels in considering would increase the overall quality of instructions and allow even those with low capacity to properly acquire and retain the new knowledge provided. Sweller also categorized the CL into three different types: the Intrinsic Load, that was responsible for handling the complexity of a given task, the Extraneous Load, a type associated with dissociated information and distractions amidst the current learning experience, and the Germanic Load, which processed the information into the LTM Schemas.

The GE was first explored by Slamecka and Graft in the year 1978, a recorded phenomena in which information is better retained through self-generation, as opposed to the more common reading methods, as a form of mnemonic or memory strategies. While the exact reasoning as to why it works is still in research, they are proven to hold effects beyond didactic approaches. The effect was commonly tested through incomplete lists of keywords, where participants were tasked to fill in the blanks, and when compared to those that were given complete lists to remember, the self-generated ones held better results. Some research points towards this effect being further enhanced through GCs- which refers to the amount of information an individual is given that limits what can be produced in such generated tasks. By limiting the possible answers through specific clues, we can more easily lead people to the correct answers, while not affecting the GE, which would overcome one of the main issues of the effect- more complex tasks require more mental effort, as researched in more recent years.

3

Related Work

Contents

In the following segment we will explore certain studies that have helped us formulate our current hypothesis for this project. These studies dove into topics not commonly investigated in the video game industry, with two of them being in completely different areas altogether. Regardless of the dissociation in fields, these studies can either be directly applied to our own area of research, or will be further studied in the context of this work.

3.1 Automated Complex Span Tasks

We have mentioned previously the relevance of CSTs in our research- it is through them that we will be able to measure the WM of our participants, after all. While this knowledge was first discovered by other studies we have touched upon before, it was through an article by Thomas S. Redick, et. al. [13] that it came to our attention the individual worth of a wide variety of approaches on the matter, such as the Operation and Reading Span Tasks, which we will discuss soon. The aim of this study was to define a way to automate the process of creating CSTs, in order for other researchers to more freely and more easily apply them to their own studies- this project being one such example of someone making use of their results.

This was possible by combining the samples of multiple testing locations, with over 6000 young adult participants performing one of three types of CSTs- the Operation, Symmetry and Reading Span CSTs we have touched upon in the previous chapter of this work, more specifically when we talked about WM capacity measures. These tasks are completely computerized, generating a random combination of trials and list lengths in each instance, which in turn can be utilized on participants, automatically scoring them after the process. As an example utilizing the Operation Span Task, which involves participants solving arithmetic equations while remembering a list of unrelated words, the system would generate the necessary words to be remembered, as well as the operations required, while also confirming the validity of the answers. After repeating the process three to seven times, participants would then be requested to click the letters they saw during the trial in the presented order on a recall grid. The process would be similar in the remaining two types, but with their respective methods- Reading Span would make use of unrelated sentences instead of arithmetic equations, and the Symmetry Span would use symmetry and spatial awareness of elements shown.

This way, they were able to test the validity and reliability of these different methods and provide the results to anyone interested in applying CSTs as a WM measure to their own research, verifying which one would best apply to the situation at hand. They were able to examine a number of new variables of these automated CSTs, such as:

- Internal Consistency, which assesses the correlation between multiple items in a test intended to measure the same construct;

- Convergent Validity, that determines how closely a test is related with to other tests that compute the same or similar concepts;
- Normative Statistics, data from a given population used to establish a baseline distribution for a score or measurement.

These results were also compared to previously conducted studies in other automated CSTs tests, data which consisted of:

- Test-retest Reliability, a measure of accuracy obtained by conducting the same test twice over a period of time to the same group of individuals;
- Construct Validity, another type of reliability measurement in which we evaluate how well a test reflects the intended construct, or the conceptualized variable;
- Criterion-related Validity, used to evaluate how well a test measures the outcome it was designed to calculate.

The results indicated that these automated CSTs show valuable psychometric properties- specifically through the high results shown in the test-retest validity, their internal consistency, convergent and discriminant construct validity, and criterion-related validity. The analysis provided also showed consistency with the expected cognitive ability of the population from the sample chosen.

3.2 Player-based Adapted Tutorials

Fortunately, our work is not the only aiming to better welcome any player into the world of the video game industry- Victor Ribeiro had too hoped to enhance the opening act to most video game experiences, the tutorial, by adapting it to different types of players. Unlike ours, however, which aims to understand the influences one's inner ability to learn has in mastering mechanics in our field, Victor Ribeiro had hoped to find a way to adapt any tutorial to any type of player. While this might seem redundant to our study, outside of the fact that it also aimed to enhance tutorials, it is important to understand other research done with similar goals in mind, and derive information that could potentially be beneficial to our studies nonetheless.

In order to achieve his goal, Victor Ribeiro's study based itself around the BrainHex Model [12]- a neurobiological typology model used to categorize gamers' personalities and their preferences associated with video games. By designing the same level multiple times, each crafted to satisfy an individual archetype while all still teaching the same concepts, Victor Ribeiro wished to see if these, when compared to a player's opposite preference, their Exception, would bear better results when matched with their corresponding personality type. While the results proved negative in most cases, it did show us an exception to the rule when it comes to the "Casual" type of players, who performed better in their own field of preference. This implies that dedicated tutorials are instead best suited for casual players, and not to those more invested in the media that would otherwise fit in specific categories of play-style. Additionally, this experiment allowed us to gather that the levels designed around players' Exceptions generally made them feel more component, due to one overcoming their shortcomings through adaptation in a scenario they are less comfortable with, thus proving more rewarding. The only outlier from this case were those whose Exception score was far too low.

3.3 Generated Video Game Tutorials

Artificial Intelligence, or AI, has always been present in video games since its conception, and growing as rampant in the more recent years. This relationship is of course a given, considering how useful it is to have AI generated content in your video games to better diversify the creation of worlds, or even responses of our NPCs, among many other utilities. But it was only more recently that someone decided to apply it to the first key component of more complex video games- the tutorial. Michael Green et. al. [14] turned their attention towards the possibility of automatically generating the learning stages of video games, through the use of AI that would complete a given level and generate the necessary information to teach players through its self-learning capabilities. This research shows that an interest has been developed in optimizing tutorials for our players, in order to best suit their needs. Being able to adapt tutorials to our players based on given parameters could revolutionize digital games as a whole, and if our research bears fruit, this approach could be then integrated to take WM levels into account as well, generating tutorials depending on the results.

Their approach branched into multiple paths- the AI could generate natural language descriptions of game rules, it could also generate instructive game levels carefully designed around critical knowledge, or even generate demonstrations of how to play a game through demonstrations. To accomplish this, the AI would structure the information gathered through the completion of levels in graphs. These graphs were based off the discovered mechanics, where each node represented an individual mechanic utilized, and connected nodes were related in some shape or form. The leaf nodes would be given a final state, be it "Win" or "Lose", where the "Win" states would be potentially turned into critical paths, which were the shortest interaction chain necessary to win the current level- that is, the recommended course of action given by the AI. Other information could be drawn from this graph, such as how frequently a given mechanic should be used. Alternatively, they also proposed that a more human-behaved AI could instead be armed with the necessary knowledge of the game rules, and made to play the game to find critical paths as well.

3.4 Summary

Thomas S. Redick, et. al aimed to research the automatizing of CSTs, in the hopes that other researchers to more freely and more easily apply them to their own studies. They achieved this by performing tests to over 6000 young adult participants, who performed one of the three most used types of CSTs: Operation, Symmetry and Reading Span. Through these results, they were able to verify the reliability of the different methods, where they differentiated in results, and where to best apply each to a given situation. The results accounted for the Internal Consistency, Convergent Validity and Normative Statistics of each experiment, data that was then added and compared to previously tested information, such as the Test-retest Reliability, the Construct Validity and the Criterion-related Validity of each approach.

On another subject, Victor Ribeiro instead focused his efforts into finding a way to adapt tutorials, much like we aim to, but instead to the type of player one was, as opposed to analyzing their psychological data. His research was done based on the BrainHex Model, where Victor created the same level multiple times and applied necessary differences to each version to better suit a given personality type, while still teaching the same fundamental mechanics. The participants were asked to play both the associated level with their player type, as well as the Exception case, or the opposite side of the spectrum, to verify if there was a major difference in results to what was supposed to be the most adapted and least adapted cases to the individual. The results proved, however, mostly negative, with the exception of the “Casual” type of player, that excelled in the levels designed around their personality- additionally, most players felt more competent after completing their Exception level.

Michael Green et. al aimed to apply Artificial Intelligence within the context of tutorials, generating video game tutorials through its use. The idea was to make an AI complete the levels ahead of time, allowing it to formulate graphs of information based on the mechanics required to complete a level, which it would use to formulate critical paths and instruct the player on how to best approach the stage. This was also hypothesized to also be possible through the use of a more human-like AI, which would be armed with the necessary rules of the game, before being tasked to complete the levels presented to it. Their approach varied into three paths, where one aimed to generate natural language descriptions of game rules, the other instructive game levels carefully designed around critical knowledge, an a final one generating demonstrations of how to play a game through demonstrations.

4

Implementation

Contents

In this chapter, we will explain the intended approach we wish to take in order to tackle the issue presented earlier in this work, as well as the environment we will utilize to run it and the methods utilized to gather the information necessary to draw a proper conclusion on the matter.

4.1 Approach

Our goal with this work is to find the best approach to tutorials, as in verifying which method is capable of better adapting to multiple levels of knowledge retention in players- their WM capacity [2]. After all, in video games we often have to learn new mechanics and apply them in complex scenarios, with our WM constantly engaged at least to some degree- the better we retain the information, the quicker and more efficiently can we apply what we have learned. But before anything else, we must determine if the gathered evidence supports our hypothesis, which aims to verify the influence of our WM capacity in a video game environment, as well as compare different approaches to tutorials in order to best adapt for said differences in players.

In order to utilize the desired WM capacity tests, the CSTs [7], not a lot needs to be changed beyond the usual approach to them, since in our situation we can simply apply the existing methodology directly to our project. The only difference it will have compared to the usual CST test relates to where exactly they will be created and applied- in this case, in the context of our Game Engine of choice. Changing its structure and rules would require us to verify its validity, which is our case is an unnecessary step to take.

As far as the approach to the tutorials that will be developed in the scope of this project, we have decided to make use of polar opposite scenarios, utilizing the common disparity between didactic and self-generated learning- or the GE [10]- in the context of a video game, often debated in the education field. While this theory has not been tested in digital games, tutorials are as much of a learning experience as any other given in a classroom, where the teacher is the game itself, and the student the player. Our brain is engaged in similar ways, and information is gathered through the same sections. The Mental Effort Theory [11] we have established before also points to a connection between the GE and our CL, or Mental Effort, which is the usage of our WM resources. As such, we believe this comparison between different teaching methodologies has unexplored potential in our field that has enough evidence to justify it.

With the previous studies we have explored before, we have the necessary backing for the reasoning behind our experiment, and will now explain our approach. As mentioned above, we aim to utilize CSTs to determine the capacity of our participants' WM [6]- we wish to test the influence of this data when applied to real scenarios, opposite to each other. The didactic tutorial approach should provide immediate results when acquiring new knowledge, but fall behind in the long run, while the GE approach has been proven to wield different results depending on the complexity of the tasks, which would imply an influence in WM levels of the testers GE. Granted, we do not expect immediate results just on the tutorial stage- especially since it has been established that complex tasks take longer to be absorbed through the GE methodology, while wielding better results in later stages of experiments. Because of that fact, we require a longer experiment than usual, with a few initial levels to further test the retention of new knowledge perceived in our participants.

This work will be developed around a Tactical Role-playing Game/Puzzle Game known as **"Nick of Time"**, genres chosen due to their more complex tactical nature and room for more complicated mechanics, which can be used to better test the capacity of our participants. Beyond it, we will make use of the Game Engine itself and a couple of surveys to collect as much information as possible, be it in real-time or otherwise, to properly investigate the synergy between these two factors- the learning approach and the WM capacity of the participating- in a video game context.

4.2 Scenario: Nick of Time

4.2.1 Concept

In order to test our hypothesis, it was necessary to create its environment first. To accomplish this, a game was designed from scratch, as that would allow us to shape it from the ground up to fit the requirements we needed for this study. This so called environment, the video-game created to house our hypothesis, was titled **"Nick of Time"**, named after one of its core mechanics and overall main appeal. Its original concept saw it becoming a 2D Tactical Role-playing Game (TRPG for short), a video game genre in which we combine the core gameplay elements of a Role-playing Game, in which players control one or multiple characters and immerse themselves in the world the game presents, with strategy elements like turn-based strategy.

The chosen platform to be utilized in the creation of the game for this study was Unity- a cross-platform game engine capable of supporting development of 2D and 3D games, popular as a starting point for game developers, and also widely used among the indie side of the industry. It has been chosen over its competitors due to the considerable amount of accessible assets online to support the creation of this project, given its popularity, and has housed well known and graphically complex titles like "Ori and the Blind Forest"[1], "Cuphead"[2] and "Subnautica"[3], only to name a few.

[1] Ori and the Blind Forest, 2015, Moon Studios, Windows/Xbox One
[2] Cuphead, 2017, Studio MDHR, Windows/Xbox One
[3] Subnautica, 2018, Unknown Worlds Entertainment, Windows

Tactical Role-playing Games are usually portrayed in grid maps, easily compared to a game of chess, where the player controls their characters like pieces in a board game with a given goal in mind, which includes but is not exclusive to eliminating all the pieces in the enemy's side of the field. One of the most iconic examples of the genre would be the "Fire Emblem"[4] franchise, where you control an army of characters in a medieval fantasy setting, notorious for its permanent death of units combined with their personification to develop an attachment to said characters. **"Nick of Time"** takes inspiration on such titles, but after careful consideration on the time available for our participants to test our game, was further defined to also have a Puzzle game aspect to it- that is to say, it turns the maps seen in Tactical Role-playing Games into a short problem to be solved by our players. In order to do so, certain elements from TRPGs had to be removed, especially given the desirable quick nature of tests like these. Notably, maps are much shorter, there is no permanent death system and the narrative structure was removed, as it simply couldn't afford to be the focus.

In exchange, the maps become shorter and thus quicker to complete. To emphasize that, **"Nick of Time"** has a limited amount of turns available for its players to complete levels, which varies depending on the current stage. The players are given a selection of characters, each with unique attributes and abilities, and a challenge in the format of a Tactical Role-playing Game's grid map, and must make a selection of which units to bring in order to solve the puzzle at hand, which in the game's current version, always means eliminating all enemies before the turn counter reaches zero. Each turn, following the standard rules from the genre associated with it, has a player side and an enemy side, which must be taken into account when formulating a strategy to complete the levels.

However, the game had to deviate somewhat from other tropes of the genre, like specific weapons often seen and associated with a given game element, in order to minimize the influence any previous experience in these types of games might have over our players. Because of this, we chose to avoid the medieval fantasy or futuristic settings seen in the more notable entries of the genre, and instead focused on a mixed fantasy scenario, where all types of creatures from different timelines can be present, thus allowing us more creativity and a broader range of options that will not be immediately understood through association.

4.2.2 Game-play

As previously explained, the game takes our players into grid maps, where they must complete a puzzle in a given amount of turns, composed by a player and an enemy side. In order to do so, the player must select only a few units from an established selection of characters, each with their own skill and at-

[4]Fire Emblem is a Tactical Role Playing Game video game franchise developed by Intelligent Systems and published by Nintendo, originally released on the Famicon, where your units are individual characters that can die and be lost permanently

tributes. The objective is to always eliminate all enemies within the time limit, which changes depending on the puzzle at hand. The key to these puzzles is often associated with the attributes of the enemies, their distribution on the map and the amount of turns given. The participants must weigh their characters' strengths and weaknesses in order to tackle each scenario. The players will have a wide variety of options to them, but the enemy will only attack enemies in range on their turn, almost as if reacting to an ambush as opposed to having the time to strategize as well.

4.2.2.A Attributes

Each character, akin to most Tactical Role-playing games, have a number of attributes associated with them that are present even in the enemies, albeit the number vary from unit to unit, and can always be checked (see Fig. 4.1). These attributes were simplified in order to keep certain aspects of the game simpler and not overwhelm our players. To start us off, we have the resource attributes, which include the Health Points and Mana Points of characters; Health Points (HP for short), determines the amount of life left in a character and how many hits they can take, which will deplete when receiving attacks, if their defenses are broken through. If this amount reaches zero, the unit is destroyed in the present map. Meanwhile, Mana Points (MP for short), is a resource utilized in a specific type of action called "Sync", which we will explain shortly. If a unit does not have the necessary amount of Mana Points to perform these actions, they will not be able to make use of them.

Then we have the offensive attributes of a unit, named Might and Magic. These attributes define how much damage a unit can deal to an enemy, physical and magical respectably. Each unit has a tag which defines which type of damage they utilize, so a physical unit will only care about Might, and vice-versa for a magical one. On the other side of the same coin, we have the defensive attributes, which dictate how much damage a unit can take of a given type: Defense for physical damage, and Resistance for magical damage. Simply put, it reduces the incoming damage of attacks of their respective type. Finally, we have the attributes which dictate the reach of a character: Movement gives you the amount of squares on the grid your character can traverse, diagonals included, while Range gives you how far the unit can attack an enemy from, diagonals included as well.

Note, however, that much like units, the map itself has its own attributes to consider, designated as "Terrain". Different terrain can influence how you see the puzzle, ranging from simple squares which cost more to traverse or are impossible to pass, fortified spots that increase one's defensive attributes, and even altitude, which will influence the damage of your attacks, given that you are higher up than an enemy. Both terrain and attributes must be taken into account before starting a map properly, as these can heavily influence who to pick for each situation.

Figure 4.1: A screenshot of a character status screen, where you can check their attributes.

4.2.2.B Actions

For each turn available to the player, all characters selected to participate in the map will be given both the opportunity to move, and a selection of actions they can take depending on what the player believes is most apt for the situation (see Fig. 4.2). These actions are what really give depth to the game, and can be summarized to three standards options often seen in other entries in the genre, as well as an unique option to this game. First, the character will have the "Move" action, which will be utilized in order to mobilize a unit to a desired space within reach, and will burn the character's movement option per turn. This action can be undone, as long as the unit has yet to act on a turn. After that, we have the "Attack" option, which is used as the main offensive option of a unit, which will allow the character to make a standard attack against an enemy with the intention of eliminating them, making use of their respective main offensive stat. Note that an attacked unit, if they still have the Health Points and range to attack their offender, will strike back, which will be important in some maps. To complete the simple options, we also have the "Wait" action, which will end the turn for the selected unit, if one does not desire to do anything with it on the player's side.

"Nick of Time"'s main feature comes from an unique action type by the name of "Sync", which holds three different options within its category, all of which make use of the Mana Points described earlier. First, we have the "Skill" action, unique to each individual character available to the player. These abilities give the player a wide array of special actions, that can vary from an attack in an area, attribute enhancements, weakened abilities to the enemies and even simple more powerful versions of regular Attacks. Then we have the "Pair Up", an action which requires an ally nearby in order to perform. This option will enhance the ability of the first selected unit with either a bonus if it targets allies, or a negative if it targets enemies. Note that this will consume the action of both allies involved, but will only use Mana Points from the ability wielder. Finally, we have the "Chain" option, which will require a minimum of three allies within reach of attacking an enemy. This action allows you to perform a sequence of attacks in a row, only allowing the enemy to potentially retaliate at the end against the first selected unit if such is possible, much like the "Attack". However, there is another caveat: akin to the "Pair Up", there will also be a buff given to all units performing the chain attack, but also a debuff. Both of these will disappear after the chain, and much like the "Pair Up", this will consume all actions of the units involved. There is a small cost of Mana Points for both selecting this action, and initiating it.

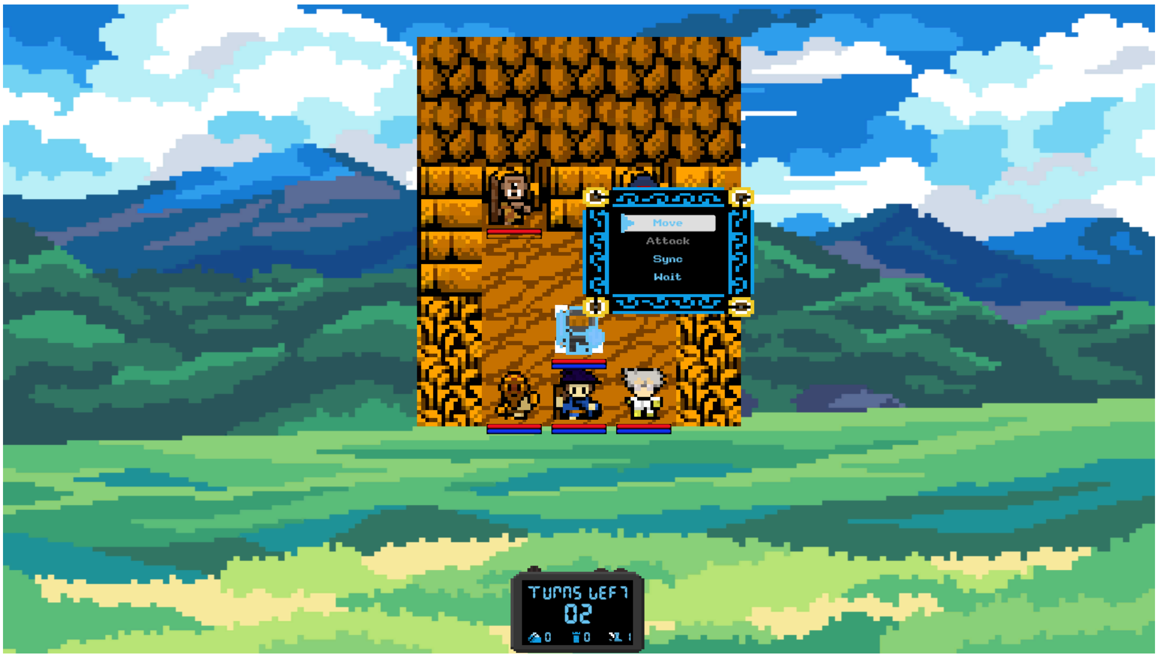

Figure 4.2: A screenshot of some action options. Note that those not available to the player will be greyed out.

These bonuses and negatives present in the "Pair Up" and "Chain" actions are not randomly selected, however, and are instead derived from a final tag present in every ally, called "Timeline". Each individual character comes from their own time period, so to speak, which influences their strategy in battle. As such, they will grant different benefits, as well as negatives, during these combination actions present in the game. Pair Ups will apply their strong suit attribute, either positively on an ally and negatively on an enemy, while the Chain will apply their strong and weak point to all allies involved, while not influencing the enemy. Tribal units boost Resistance and lower Defense, Medieval units boost Magic and lower Might, Modern Units boost Defense and lower Magic, and finally Future units boost Might and lower Resistance.

4.2.3 Architecture

Moving onto the environment itself, Nick of Time's introduction takes the form of a CST test, more specifically a version of the Operation Span Task [13] which was already scientifically proven to work in other aforementioned studies. This version, however, is fully implemented in the game itself in a digital format. In order to avoid potential incoherence with its pen and paper counterpart, the test was kept mostly accurate to its original form, including the timing between the arithmetic questions (Fig. 4.3) and the sequence input of letters (Fig. 4.4) that constitute an Operation Span Task. In this case, we opted to do a sequence of 6 operations and 6 letters to remember, the tests akin to this one choosing from 4 to 8. While not directly linked to the rest of the gameplay, this section will be relevant in the data gathering, which we will detail soon. Upon finishing the test, our participants will then be split in the tutorials.

Figure 4.3: An example of an arithmetic operation in the CST test.

Figure 4.4: An example of the sequence input in the CST test.

The tutorials we have mentioned, the GE tutorial and the didactic tutorial, have a few select differences when compared to regular levels: they lack a turn limit, allowing our participants to have as much time as they desire to experiment in the training area before they jump onto the challenges, and players will be able to revisit the character selection screen as many times as they please without having to restart the level (see the example in Fig. 4.5). These tutorials will share the same map and dummy enemies, their main difference lying instead in how new information is provided to the players. Our testers will be divided between a GE focused tutorial and a more didactic one, where the GE tutorial will only give a minimal amount of information, thus only explaining the very basic of each mechanic and letting the players complete the missing pieces by themselves, while the didactic tutorial will portray the same elements extensively, guiding the players step by step. What we aim for here is to in one case give the players total freedom on how they learn, while in the other scenario they are given everything they must know. To drive the point across, the GE players will always be able to skip ahead of each tutorial section, while the didactic participants must first follow the lesson before giving the option to continue.

The difference will be most notable when the core mechanic of the game is revealed: the “Sync” action. Characters will each have their unique set of abilities, and through teamwork, the player will be able to combine these units and enhance their techniques in order to beat increasingly more difficult challenges that the individual characters wouldn’t be able to on their own- or at least, not as efficiently. To further exemplify the differences in tutorials, the GE method will only mention how to access these skills and what they are, while the didactic one will detail more information and take players step by step through each subcategory of the action type.

Figure 4.5: A screenshot of the Generation Effect tutorial, where the player skip each section at their leisure, as well as an example of the character selection.

After the tutorials, all participants will converge into a sequence of three levels (Fig. 4.6, Fig. 4.7 and Fig. 4.8), with the difficulty increasing after each one. Each individual map will have its own layout, both of terrain structure and enemy positioning, as well as the turn limit previously mentioned. Part of the gameplay loop includes studying these changes and adapting, which is an important step on what we aim to study in our research. Common to all maps, however, is how it starts: with the character selection menu. Since players must only pick a few of the units available to them, they are given the option to view the map and inspect enemies before selecting said units, organizing them in the starting positions available to them. Once the player is ready, they can begin the challenge when they see fit, and restart if they believe they have done something wrong whenever they please. All the maps available will start with the player's side, moving onto the enemy turn once all units have finished their actions, and then back onto the player's turn until the counter at the bottom reaches zero or a player manages to clear out all enemies present in the map.

Figure 4.6: A screenshot of Level 1, a long map with high damage enemies. The player must watch out for the enemy round, since they can defeat their units easily.

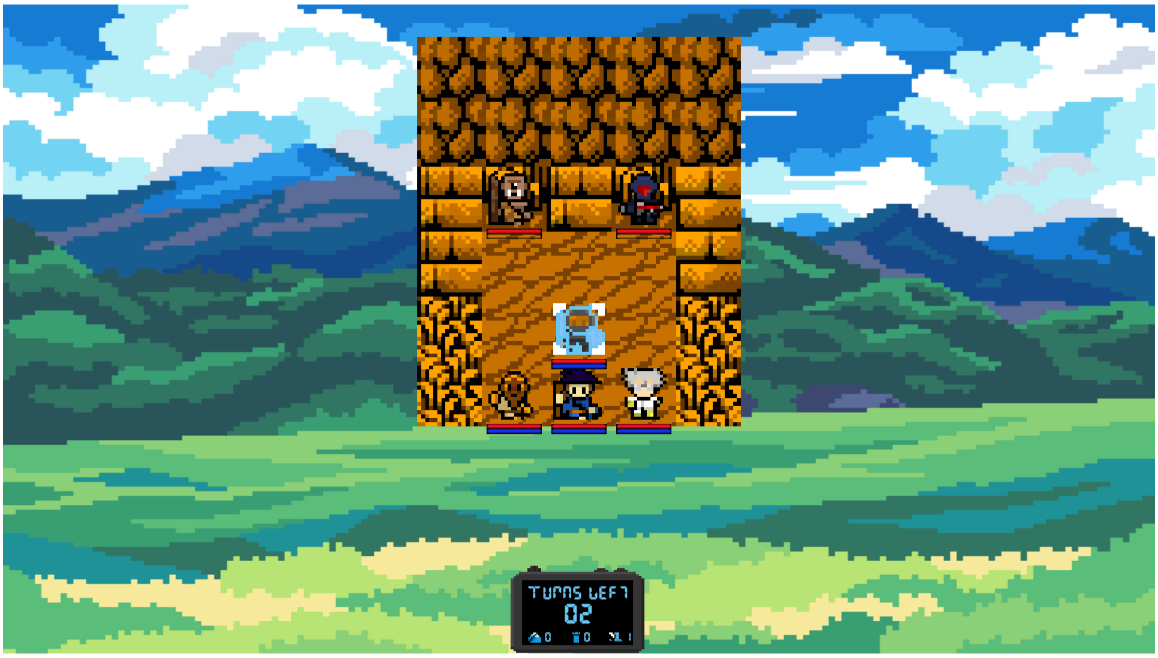

Figure 4.7: A screenshot of Level 2, a shorter map with two very strong units. They each can defeat any of player unit on their own, but have exploitable weaknesses using the ”Chain” action.

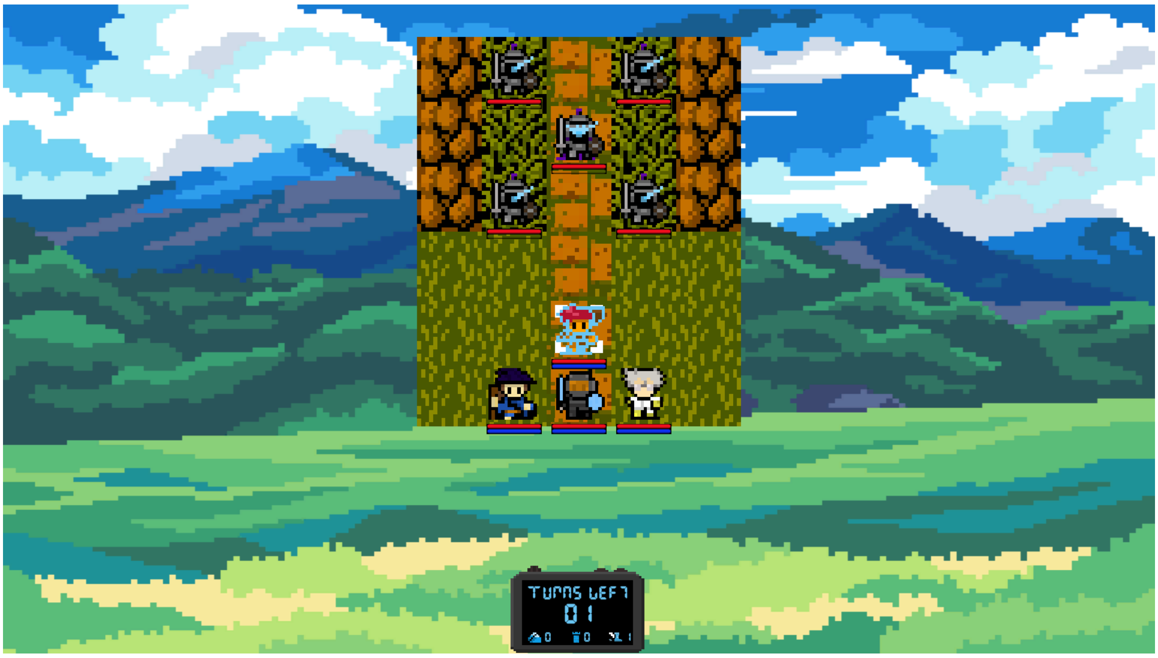

Figure 4.8: A screenshot of Level 3, where different kinds of enemies are present. The player must make use of the units’ unique abilities to overcome this challenge, as well as manipulating who they attack on their go.

4.3 Collected Data

Our collected data will be separated in three different groups: one will focus on demographic questions, the other will be extracted directly from scripts during playtime of Nick of Time and the final one will consist of feedback questions regarding the workload felt. The demographic questions will consist of questions in relation to age, gender, how often our participants play video-games on their daily lives and how familiar they are the genres portrayed in Nick of Time. These will be utilized to avoid making incorrect conclusions with the data collected, as we do not want certain aspects outside of our experiment to influence the results. For example, someone with little experience in video games might have different results than a participant who actively plays them and also prefers the genres of our video-game, and we must watch out for such variables.

Once in the game "Nick of Time" itself, the scripts on the background will take care of retrieving the data we desire. These scripts can be separated in the three main phases of the game: the CST test, the tutorials and the levels themselves. During the CST test, the only relevant component to be collected are the results of the test itself, which will be crucial to be able to draw conclusions from the rest of the experiment. Note that the arithmetic operations are meant to be used only to validate the attempt, and as such any participants who do not meet a quota of at least one operation correctly answered will have their attempt nullified, regardless if they get all the sequences right. This is utilized to make sure participants don't just focus on one side of the test, as the real quantified data will be the sequence to remember. In the tutorials, our scripts will only collect the time taken to finish them, which should give us an idea of how much time was dedicated in each type to learning the mechanics presented. Finally, during the three following levels, we will collecting a bigger sample of data than in previous phases. We will be keeping track of how much time a player spends in a given level per try, the amount of tries they take to complete it, which units were selected and which actions they pick. Note that these levels were purposely made moderately difficult, since easier levels would most likely not allow us to gather any meaningful information. All of this data will allow us to comprehend a multitude of things, be it a player's struggle in understanding the challenge presented, which mechanics they remember to utilize, and as a bonus, even data that could indicate a type of strategy that may be over-tuned and over-used to complete the levels. Thanks to this data, we will also be able to paint a picture of their learning curve throughout the levels and how it correlates to both their WM and which tutorial type they were given beforehand. We theorize that those that took the didactic tutorial will have an advantage in the first stages, since they were given all the information from the get go, compared to those in the GE tutorial, which were taught from their experience only. However, the latter should see improvements in the long run. That being said, those with lower WM levels should still be struggling more than those with higher capacity, regardless of which tutorial taken. The gathered data will be uploaded into a text file, which we will be requesting from our participants after they conclude the game.

To finalize the data collected, there will be a second part of the survey in which we will inquire some feedback on the video-game our participants had just finished. Most of these questions will take on the form of a simplified version of NASA's Task Load Index [18], or NASA TLX for short, which is a assessment tool used to rate perceived workload in a given task, developed by the Human Performance Group at NASA's Ames Research Center. This part of the survey will help us verify the quality of our video-game environment, as the workload refers to the amount of WM resources used [19]- that is to say, the higher the workload, the more resources the participants used (from their perspective, in this case).

4.4 Summary

This project's objective is to determine an approach to adapting tutorials to multiple levels of WM, as we believe understanding these levels are key to allowing our players to better retain new skills taught to them in video-game environments. To accomplish this, we intend on utilizing a measuring test referred to as CST, so we may inquire the WM levels present in our participants and how they behave under different teaching environments- in our case, we picked two polar opposite approaches, making use of the GE exploratory approach and a more didactic extensive method, a popular comparison often discussed on the topic of education.

The environment chosen to test our hypothesis was a game specifically made in the context of this work, a 2D Tactical Role-playing Game mixed with a Puzzle Game by the name of "Nick of Time", utilizing Unity as our Game Engine of choice. Our game takes the grid-like maps of other titles of the Tactical Role-playing Game genre and gives it a twist, a limited amount of turns in which the players must complete the challenge presented. Each map is filled with different terrains and enemies, which the player must analyze and pick from a selection of units which ones are best suited to handle the task at hand, by taking their potential actions, unique abilities and their attributes into consideration.

The game will start participants with the CST test, in order to quantify their WM, before dividing them between the two types of tutorials: the GE and the didactic environments. Once a player finishes the tutorial, regardless of which tutorial they were given, they will all move onto a sequence of three levels, whose difficulty increases the further one goes, made to test their knowledge of the acquired skills. All of these sections will contain scripts that will gather a multitude of information, from the results of the test to the time and tries taken per map. Additionally, we will add two sections of a survey: one for demographic questions, to ensure our data is not skewed by outside factors, and one for a simplified NASA TLX version, in order to verify the quality of the learning environment.

5

Evaluation

Contents

In this chapter, we will describe the procedure taken to run this experiment, detailing each of the steps and what purpose they serve. Afterwards, we will detail the results gathered through said procedure, as well as analyze the data collected in order to verify our hypothesis' validity.

5.1 Procedure

Our procedure can be divided in three main segments, all of which were done online, which followed the setup by the "Collected Data" section; an initial demographic section, the short video-game Nick of Time in which the main tests will be setup and a follow-up section with subjective questions regarding workload. To accomplish this specific structure, we decided to make use of a Google Forms (see the Appendix for details), where we present all three of these stages. Our first page consist mainly of a short explanation of our study and its intended purpose, as well as serving as a consent form for the participation of the remaining sections. The participants that chose to continue were presented with the demographic questions, previously explained in the last chapter, and only once it is concluded will they gain access to the second section, where a short explanation of Nick of Time, together with a drive link to download it, are presented.

The player will then jump onto the game, which will begin the CST test. Considering the nature of the test, we provided our participants with an explanation of how it operates, including a free trial of a short version of it, where the participant can retry as many times to get used to the short timers. Only the player feels ready, they can begin the real trial. Once it is finished, the game will take the player to one of two tutorials, the GE version or the didactic one, which will follow the behavior explained previously. Following the tutorial, the players will be moved onto the first of three levels, and will be tasked to complete as many of them as possible. Considering the difficult nature of the challenges presented, they were given the option to back out once they felt like giving up. While we understand that doing so might persuade players to throw in the towel early, we couldn't afford to simply frustrate the participants beyond an unnecessary level. Regardless, we believe their attempts and early surrender will give us valuable information regarding the tutorials and their CST levels.

Once the player concludes the video-game, be it by finishing the last level or giving up somewhere through the levels, they must then return to the Google Forms from earlier. There, they will be requested to submit a file generated by the game, containing all their collected data throughout the video-game experience in Nick of Time, as well as provide feedback on the tutorial they had received. After doing so, the final section will be available, containing the simplified version of the NASA TLX, constituted by six simple questions regarding the workload felt during the experiment, regarding: Frustration, Temporal Demand, Physical Demand, Mental Demand, Performance and Effort, which must be scored from 1 to

10, as opposed to 1 to 20, due to Google Forms' restriction on scale questions. This section will conclude the participant's task, and the overall structure we followed throughout the experiment described here can be seen in the figure down below (Fig. 5.1), for the sake of clarity.

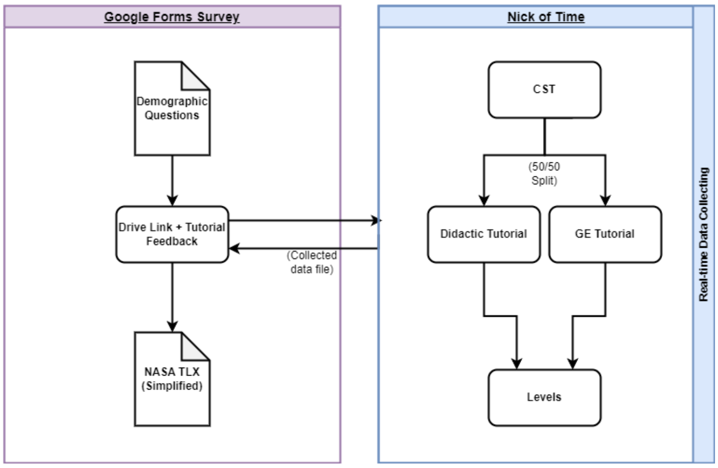

Figure 5.1: The procedure utilized in this research.

5.2 Sample

To start off, we ran a short pilot in order to ensure the process could be done without any major hiccups that were not considered or tested previously. The pilot, much like the real run of the experiment, was done online through the use of the aforementioned Google Forms and Google Drive link. We asked 5 participants to help solidify the experiment, and thanks to it, we were able to discern some significant bugs and, more importantly, the difficulty of the levels. The issue lied in the order in which the levels were presented, and as such they were correctly shuffled to better suit their difficulty. Any bugs found during this short pilot were also taken care of.

The sample itself, however, only amounted to 17 new participants after the original test run, and as such, does not possess enough data to draw any proper conclusion. Given this fact, we will treat the results presented in the remainder of this document as a pilot study, helping future implementations of this theory to have a base defined for them. That being said, we will still be analyzing the collected data and verifying our hypothesis, based on the limited data we gathered. The data presented was extracted from the Google Sheet generated by the Google Forms and another extracted from the data generated by the game and delivered by our participants in the questionnaire.

To start us off, we will analyze the demographic section. From our sample, 14 (82,4%) participants identify themselves as male, with only 3 (17,6%) identifying as female (see Fig. 5.2), while the age average collected rounds to 25, with ages ranging from 21 to 34 years old (see Fig. 5.3).

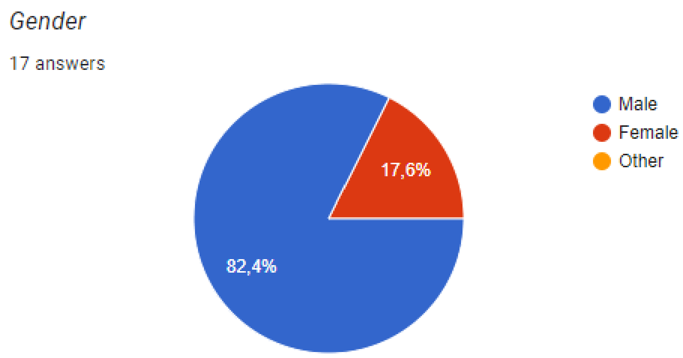

Figure 5.2: Pie chart with the gender information of our 17 participants - 14 male and 3 female.

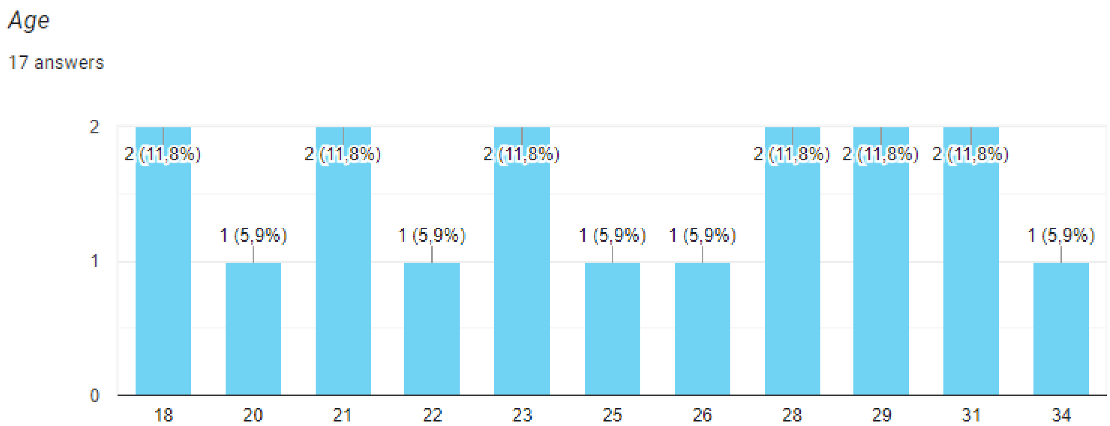

Figure 5.3: Bar chart with the different participants' age.

Out of the 17 participants, all but one confirmed they make time in their schedule to play video-games regularly, with the outlier only playing them socially (see Fig. 5.4). As far as the genres presented are concerned, a majority of our participants enjoy (41,2%) or favor (35,3%) Tactical Role-playing games, with the remaining percentage either not having a formed opinion (11,8%) or not favoring them (11,8%) (see Fig. 5.5). Puzzle games ended up having more defined opinions, with close to half enjoying them (47,1%), but about a quarter not favoring the genre (23,5%), the remaining participants favoring Puzzle games as one of their favorites (see Fig. 5.6).

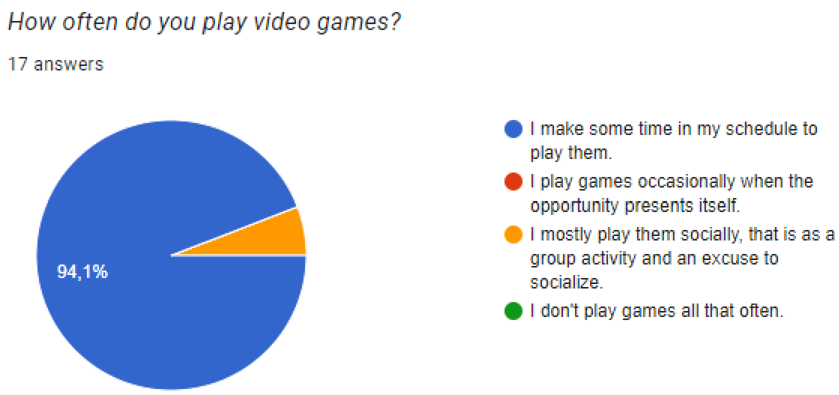

Figure 5.4: Pie chart with the participants' answers when asked how often they play video-games.

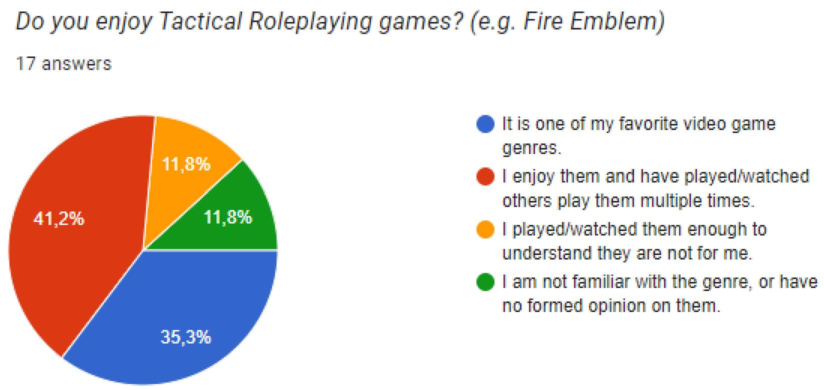

Figure 5.5: Pie chart with the participants' answers when asked how familiar they are with Tactical Role-playing Games.

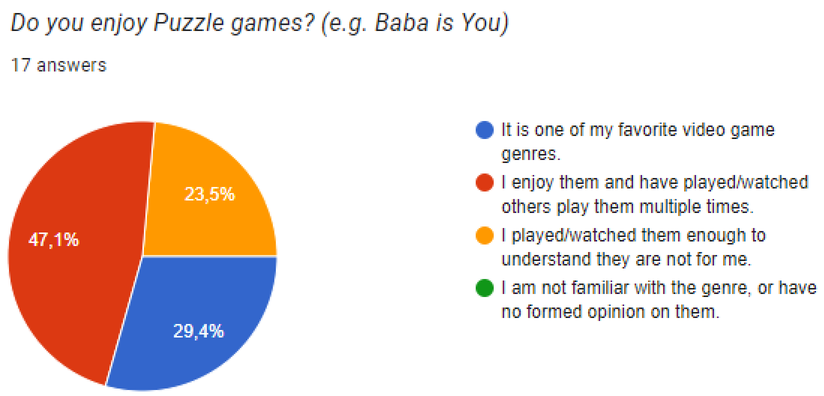

Figure 5.6: Pie chart with the participants' answers when asked how familiar they are with Puzzle Games.

Moving onto the data collected through the scripts in Nick of Time, we can verify that our CST results have a median score of 3.5 on a scale of 1 to 6, meaning that it lands somewhere between 3 to 4 correct letter sequences. However, 3 out of the 17 participants did not get enough inquiries correct to qualify, be it from a lack of correct arithmetic operations or in general, reducing our real data sample to 14. Out of those 14 participants, 9 ended with the GE tutorial and averaged 7,8 minutes spent in it, the remaining 5 getting the didactic tutorial with an average of 16 minutes spent in it. Those with above average scores in the CST spent more time in their tutorials, around 11,3 minutes, while those with below average scores only spent 7 minutes in their respective tutorials.

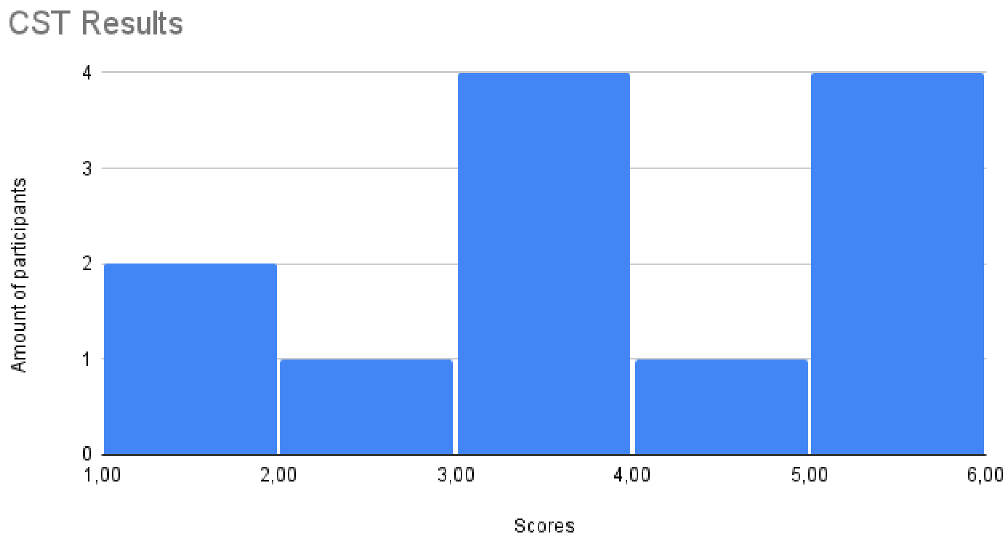

Figure 5.7: Bar chart with the different participants' CST scores.

On the levels, participants needed an average of 3 tries to beat the first level, with 5 participants giving up on it and only one having a low CST score. On the second level the tries required increased to an average of 8 with 2 more participants giving up, the remaining ones moving onto the final level with an average of 11 tries and 3 more quits. This means that only 4 participants truly concluded Nick of Time, with only 1 of them taking the didactic tutorial in the beginning, and only 1 of them having a low CST score, albeit barely with a 50%. It is also noteworthy to mention that all of those from the GE tutorial that finished were experienced players in the genre, while the only player that beat Nick of Time from the didactic tutorial only had some experience with Tactical Role-playing Games, but did not enjoy Puzzle games. Out of the didactic tutorial participants, 3 quit right from the first level, meaning the results from that tutorial were cut shorter than expected.

When looking at the second part of the Google Forms survey, we can note that the didactic tutorial scored an average of 6 out of 10 from the subjective opinion of our participants, while the GE tutorial scored 5,5 out of 10. In the NASA TLX section, Mental Demand averaged on 7,6 out of 10, Physical Demand only averaged on 3 out of 10, Temporal Demand scored 5,9 out of 10, Performance on 5,3 out of 10, Effort with 7,5 out of 10 and finally Frustration with 6,5 out of 10. That means that on total the task our participants performed was rated a 35,8 out of 60, on average.

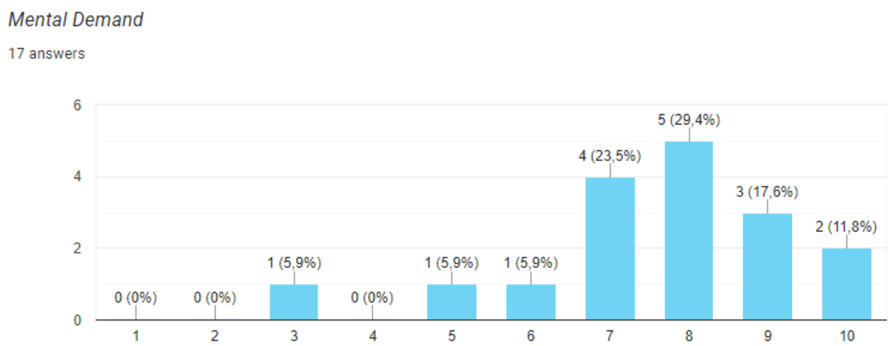

Figure 5.8: Bar chart with the scores attributed in the NASA TLX's Mental Demand section.

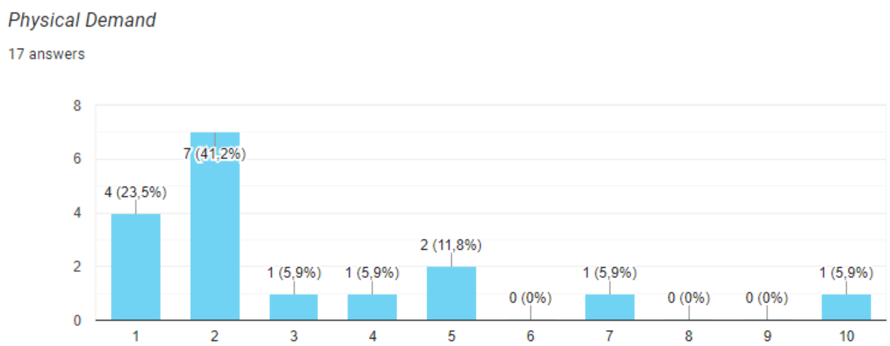

Figure 5.9: Bar chart with the scores attributed in the NASA TLX's Physical Demand section.

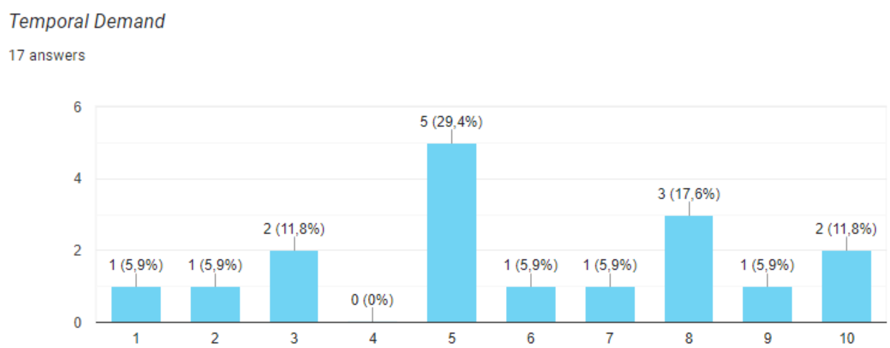

Figure 5.10: Bar chart with the scores attributed in the NASA TLX's Temporal Demand section.

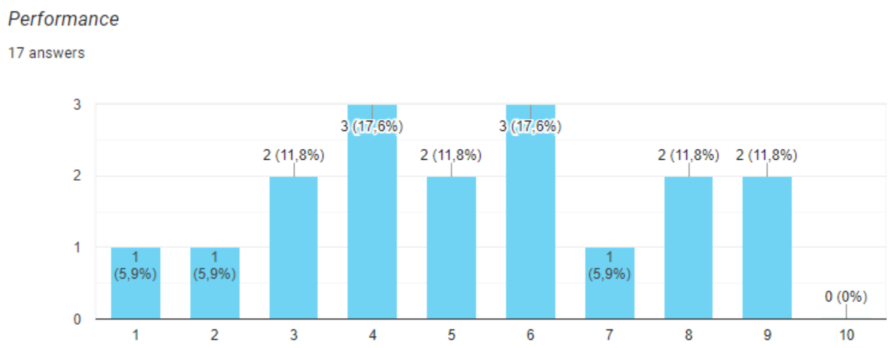

Figure 5.11: Bar chart with the scores attributed in the NASA TLX's Performance section.

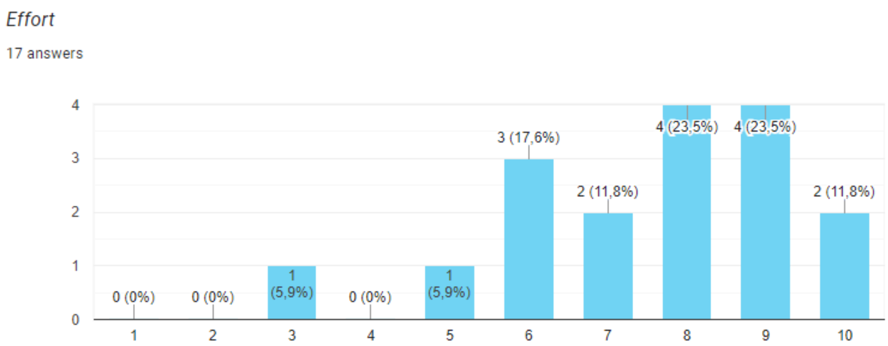

Figure 5.12: Bar chart with the scores attributed in the NASA TLX's Effort section.

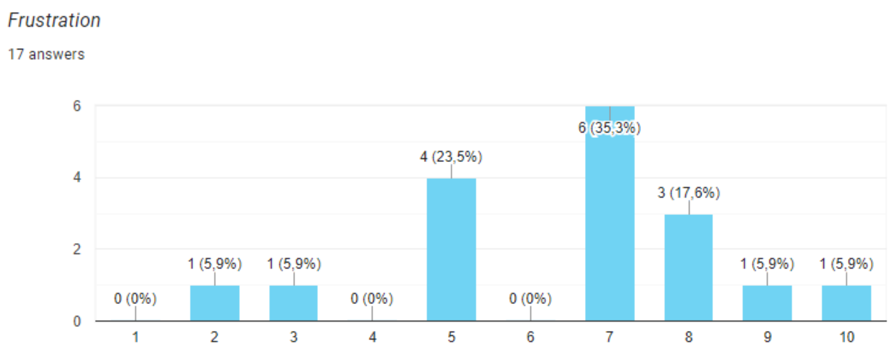

Figure 5.13: Bar chart with the scores attributed in the NASA TLX's Frustration section.

5.3 Results

Before we theorize with our results, it must be clarified again that the sample collected is not large enough to determine anything concrete, and as such we will analyze the data from the point of a view of a pilot instead. Summarizing, we had hypothesized that participants with lower WM levels, i.e., that scored lower in the CST test, would tend to perform better in a didactic environment, since it would guide them slowly through what they had to learn, while those with higher levels would perform better in the GE tutorial, since their higher capacity would allow them to understand mechanics quicker with the freedom provided from a more exploratory approach.

First, we can verify that participants in the didactic tutorial stayed in it for considerably more time than the participants in the GE tutorial stayed in their respective environment. That is to be expected, considering how the didactic approach takes its time to guide the player through the mechanics in detail, but the fact that the average time ended up being almost doubled when compared to the GE approach may indicate that when a player is given the chance to dictate their own rhythm, it will lead to them potentially investing less time in learning, or trimming out information they might already be familiar with, as we will soon discuss.

Most of those that concluded all the levels originated from the GE tutorial, and managed to complete said levels in less turns than the one participant from the didactic side. Additionally, we can verify that most of the participants that finished Nick of Time were, indeed, those with higher WM presented. In fact, the only participant from the didactic tutorial that finished the game had an average to low score in the CST, meaning that the present data does not contradict our original theory, and those with higher WM do perform better in an exploratory approach like the GE teaching method, while those with lower levels perform better with didactic teaching methods.

Another point to bring up is how those that beat all the stages in the GE tutorial were players that favored the genres present in Nick of Time, while the one participant on the didactic side only enjoyed Tactical Role-playing Games, but not Puzzles. The GE approach does favor experienced participants in video-games, even more so those with experience in the genres selected, due to it allowing players to select which areas to explore and dedicate time to. Meanwhile, those that lack in experience might do better in the didactic tutorial, since it provides more information on topics the experienced player might already be familiar with. This could imply that different approaches favor experience instead of WM levels.

Unfortunately, we do not have enough data to support either theory just yet, even if the data prove favorable thus far. There are too many dimensions to consider, most of which share conditions and are interconnected, which makes it difficult to distinguish the real cause with a small sample. On that note, it should be mentioned that both tutorials scored similarly on the survey, so no concrete evidence could be derived from that data thus far. This may be due to the questions' vague nature that could lead those answering it to be rating it from a critical stand point, and not from their personal preference. The average score the task received through the simplified NASA TLX was relatively middling, meaning that the video-game environment was successful enough not to overwhelm our players' memory capacity [19].

5.4 Summary

This test's environment is divided in three stages, shared through a Google Forms format: a demographic section, the short video-game Nick of Time and a follow-up section with subjective questions regarding workload, in that order. The demographic section allows us to first identify our sample, and in the second section of the survey our participants will be explained what the game will entail, as well as be provided a link to download it. The game is then structured in three parts as well, starting with a CST test, followed by one of two tutorials (GE or didactic) and once it is done, a sequence of three levels of increasing difficulty. Only once the game is attempted past its tutorial, will players be able to submit the generated file and access the final segment of the Google Forms survey, constituted by a grading question for the tutorial and a simplified version of the NASA TLX.

Though unfortunate, it was not possible to gather sufficient data through the online method described, with only 17 participants and an initial 5 for a pilot, and as such we considered this trial as a pilot to a potential future follow-up. Still, by analyzing the data gathered, we could conjure certain theories of where the sample was leading to. First of all, we could note that participants in the didactic tutorial took almost double the amount of time, on average, to conclude it when compared to the GE tutorial. This may lead us to believe that players with the freedom to learn at their own rhythm will choose to end it much earlier those without the option.

Since players were given the option to back out during the levels, only a select few participants were able to finish all three levels presented, and more interestingly, most of those that finished were GE participants. Interestingly enough, those with more experience in the genres presented saw better results on the GE approach, while those without experience did better on the didactic tutorial, at least with the data available.

6

Conclusion

Contents

In this final chapter, we will condense all the important points mentioned across this study, summarizing it all in a digestible format to conclude it. Additionally, we will also address the limitations found throughout this study and remark potential flaws in its design, as well as suggestions on how to correct them for future implementations of our presented theory.

6.1 Summary of the Work

Video games grow more and more complex as years pass by, especially considering how quickly it has done so. However, it creates a problem derived from such an exponential growth: more complex games bring more complex mechanics, and it creates a need for good tutorial environments that can properly convey the information in a clear and easy to retain way. Video games do not take into account the different learning growths, and that pushes part of the player-base away- but we hypothesize that the key to solving this problem lies outside our industry, and instead connects to our memory, which is the one who is truly responsible for gathering new knowledge.

To accomplish this, we imported a variety of concepts and theories from cognitive psychology in order to properly understand how to analyze it in players and derive conclusions. First, we had to comprehend how it is believed our memory is structured, and learnt that it can be divided in two major components: the LTM and the STM [2]. The LTM is responsible for storing long-term information and skills, with a potentially unlimited capacity and for a long time. The more new information is trained, the longer it tends to stay in our memory. Meanwhile the STM, also known as WM, is responsible for holding information temporarily with a limited capacity, which varies from person to person, and is responsible for a lot of our decision making, and is the first step towards the retention of new knowledge. The bigger our WM, the bigger the amount of information we can process at a time.

The key to our theory lied in the conversion between the two, and if we can optimize it in our tutorials regardless of our players' WM capacity. To test it, we needed to measure this capacity, and watch how different levels behaved in different learning environments, in order to understand if there was a best approach to memory retention. For us to measure the WM capacity, we made use of the CST test [6], exercises that mix memory tasks, such as remembering a sequence of objects, with interleaved secondary processing tasks, like judging the correctness of equations, with the objective of targeting different brain regions. In our case, we made use of the Operation Span Task [7], that makes its participants remember an increasingly larger sequence of consonants while mixing in arithmetic operations between each sequence. We then turned our attention to the tutorial environments we wished to study. That is when we landed on two opposite approaches, often compared in the education field, the didactic and GE [10] approach. The latter, not yet applied to the video-game industry to our knowledge, made use of incomplete information to prompt students to complete the missing pieces, and thus retain the information better. We believed that such an exploratory approach, where the player must find out the details of each mechanic on their own, could prove to be a useful comparison to a relatively common tutorial like the didactic one, where detailed information is slowly given to the players.

Upon studying these approaches, we believed that the didactic tutorial approach should provide immediate results when acquiring new knowledge, but fall behind in the long run, while the GE approach has been proven to wield different results depending on the complexity of the tasks, which would imply an influence in WM levels of the testers GE, with the potential to provide better results with higher levels of WM. Studying its influence in both approaches was set to be our goal, and in order to do so, we required a testing environment for our hypothesis- thus, we created a video-game for the purposes of this work by the name of "Nick of Time", a Tactical Role-playing game combined with a Puzzle game with an unique approach to both genres and a moderately difficult learning curve, constructed that way to ensure the differences between the learning methods and different WM levels would be noticeable in a shorter experiment. The game was structured in three main sections: first, the CST test, which we converted into an introduction to the game itself, then one of the two tutorials selected at random, where the players were taught how to play the game using different methodologies, and finally a sequence of three levels in increasing difficulty.

With the blueprint set, we then decided on how to gather the necessary data for our research. First, we needed to get the demographic information of our sample, to ensure any outside influence such as experience in games, familiarity with the genres and age would be taken into account when analyzing the information. Then, we set up in the game itself scripts that would collect data on the players' performance in each section, from collecting their results on the CST test, to which tutorial they got and how long they spent on it, to how many turns one took to complete a level and which actions and units were selected. Additionally, we decided we should reinforce our data with a subjective approach, making use of a simplified NASA TLX version [18] to measure the workload felt by our participants.

Our procedure was then separated in the same three stages, all setup in a Google Forms, where participants were first asked to fill out the demographic questions, then be given a link to download and play "Nick of Time", as well as deliver the resulting file generated by the game, and finally answer a section with an additional feedback question regarding the tutorial they got as well as the simplified NASA TLX. First, we ran a pilot to correct any major flaws in our setup, and then proceeded with the real test run. Unfortunately, we did not manage to gather enough participants in the remaining time, and thus analyzed our data as a pilot run as opposed to the full procedure.

From the gathered results, we could note that participants in the didactic tutorial took almost double the amount of time, on average, to conclude it when compared to the GE tutorial. This could mean that when given the chance, players will dictate their own rhythm and conclude the tutorial earlier. Since players were given the option to back out during the levels, only a select few participants were able to finish all three levels presented, but interestingly, most of those that finished were GE participants. On top of that, those with more experience in the genres presented saw better results on the GE approach, potentially due to the fact that its exploratory approach allows them to be selective on what needs to be learned, while those without experience did better on the didactic tutorial, perhaps thanks to how it teaches things slowly and in detail. While not conclusive with the amount of data collected, the results gathered do not contradict our hypothesis thus far, and instead built upon what we believed to be true. Additionally, the average score the task received through the simplified NASA TLX was relatively middling, meaning that the video-game environment was successful enough not to overwhelm our players' memory capacity, and thus further validated [19].

6.2 Limitations and Future Work

Ideally, a future work should be attempted with a much larger sample of players, as our test run was not significant enough in size to make conclusions. We believe that the main cause of the low amount of participants might be due to its time-consuming nature, taking from 30 to 60 minutes to undertake. When doing it online, a test run should take less time as to avoid any participants being dissuaded from taking it, an issue that is less prevalent on site.

Unfortunately, we do not believe shortening the game would be possible, as something of this genre of video-games takes time to explain their mechanics, and simplifying it further would make it considerably easier to understand and complete, which would make the data collected overall higher than normal. It could be possible to lower the overall difficulty of some of the levels, especially the last one, as to avoid so many of the participants quitting earlier, since the difficulty itself might have lost us some potential testers. There were also quite a few complaints in regards to the controls chosen for the game, and with more time, we would have liked to convert it to mouse only, or simply change the setup controls to something more intuitive, especially since issues with the placement of keys in foreign keyboards were felt by some of the participants.

Bibliography

[1] White, M. M. (2014). Learn to play: Designing tutorials for video games. CRC Press.

[2] Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In Psychology of learning and motivation (Vol. 2, pp. 89-195). Academic Press.

[3] Clopath C. (2012). Synaptic consolidation: an approach to long-term learning. Cognitive neurodynamics, 6(3), 251–257.

[4] Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. Journal of neurology, neurosurgery, and psychiatry, 20(1), 11.

[5] Baddeley, A. D., & Hitch, G. (2001). Working memory in perspective. Psychology Press.

[6] Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. Journal of verbal learning and verbal behavior, 19(4), 450-466.

[7] Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent?. Journal of memory and language, 28(2), 127-154.

[8] Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63(2), 81–97.

[9] Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: an individual differences approach. Journal of experimental psychology: General, 125(1), 4.

[10] Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. Journal of experimental Psychology: Human learning and Memory, 4(6), 592.

[11] Tyler, S. W., Hertel, P. T., McCallum, M. C., & Ellis, H. C. (1979). Cognitive effort and memory. Journal of Experimental Psychology: Human Learning and Memory, 5(6), 607.

[12] Nacke, L. E., Bateman, C., & Mandryk, R. L. (2014). BrainHex: A neurobiological gamer typology survey. Entertainment computing, 5(1), 55-62.

[13] Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. European Journal of Psychological Assessment, 28(3), 164.

[14] Green, M. C., Khalifa, A., Barros, G. A., & Togellius, J. (2017, September). " Press Space to Fire": Automatic Video Game Tutorial Generation. In Thirteenth Artificial Intelligence and Interactive Digital Entertainment Conference.

[15] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive science, 12(2), 257-285.

[16] Hart, S. G. & Staveland, L. E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) Human Mental Workload. Amsterdam: North Holland Press.

[17] Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach. Journal of educational psychology, 84(4), 429.

[18] Hart, S. G. (1986). NASA task load index (TLX).

[19] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive science, 12(2), 257-285.

A

Questionnaire

Nick of Time

Welcome,
This survey and, by extension the game, are part of a Master's dissertation from Instituto Superior Técnico in Portugal and as such, any and all participants who give us their time will be helping us gather information. For that, we are most grateful. As a far warning from the start, this can be a bit time consuming, ranging with participants taking from 30 to 60 minutes, in some cases even more. Rest assured, however, that you may invest as much time as you desire, and still contribute all the same.

Given the nature of such studies, we humbly request the utmost sincerity in your answers, as these will be utilized in the aforementioned dissertation. Please, follow the instructions set in this survey, as it will guide you through the process in a hopefully smooth experience.

This questionnaire will first ask some demographic questions, in order for us to identify our sample of participants. Following that, we will instruct you on how to initialize and play a short game called "Nick of Time", and only once it is completed we will request of you to answer the following segment of the survey, entailing the workload of the experience and data collected throughout it.

Nick of Time is a Puzzle/Tactics Roleplaying Game created for the purposes of a thesis, and is compatible with Windows and MacOS.

Before we get started, we would like to remind you that:
- The participation in this process is entirely voluntary, and as such you are always given the option to back out if you so desire.
- You have the right to ask any questions regarding this experiment at any time through the following email: nickoftime.thesis@gmail.com.
- Your participation does not entail any physical or psychological risks.
- This process is mostly anonymous- however, due to the nature of Google Forms file submission system, it will utilize the associated Google account to name the file received. If you desire to stay complety anonymous, please utilize a dummy account or other such means. Regardless, we ensure you the data will not be made public or be directly linked towards you during our experiment or otherwise, and will stay within the domain of said study.

By proceeding with this survey, you are giving us your consent to the points mentioned above.

Demographic

Age *

A sua resposta

Gender *

- ◯ Male
- ◯ Female
- ◯ Other

How often do you play video games? *

- ◯ I make some time in my schedule to play them.
- ◯ I play games occasionally when the opportunity presents itself.
- ◯ I mostly play them socially, that is as a group activity and an excuse to socialize.
- ◯ I don't play games all that often.

Do you enjoy Tactical Roleplaying games? (e.g. Fire Emblem) *

- ◯ It is one of my favorite video game genres.
- ◯ I enjoy them and have played/watched others play them multiple times.
- ◯ I played/watched them enough to understand they are not for me.
- ◯ I am not familiar with the genre, or have no formed opinion on them.

Do you enjoy Puzzle games? (e.g. Baba is You) *

- ◯ It is one of my favorite video game genres.
- ◯ I enjoy them and have played/watched others play them multiple times.
- ◯ I played/watched them enough to understand they are not for me.
- ◯ I am not familiar with the genre, or have no formed opinion on them.

Game: Nick of Time

Thank you for filling out the previous questions. From this point onward, we request you play the game Nick of Time before continuing with the rest of the questionnaire. It will be available for download shortly in this section, and you will be able to play it simply by executing the Unity executable file present in the main folder (Nick of Time.exe).

Before you start, note that the game will be challenging. That being said, you can leave the levels at any point, and deliver the file as is, if you find yourself unable to proceed. Do try as much as you can first, however. Please, acquire it from the following drive link: https://drive.google.com/file/d/18D2TgLo0ZZ2gnrLQMzXC6H2MoSq3OJsR/view?usp=share_link

As previously stated, Nick of Time is a Tactics Roleplaying game mixed in with a Puzzle game, designed as part of this study. It will start with a type of test related to our studies, followed by a transition into the game's tutorial proper. Once you have finished the tutorial you will then get to play through a series of short levels, designed to test what you have previously learned.

We request you complete this short game in a single go, as doing it in parts may alter the intended results of the experiment. Likewise, we ask our participants to not use outside help, may it be other people, a notebook or even a calculator. We thank you for your understanding.

Here are the **base controls**, so you may find yourself more ready for the game ahead of time. Note that we use the keyboard exclusively for the game, as there are bugs currently present when using the mouse:
Arrows - Movement
Z - Confirm/Select
X - Cancel
C - Inspect
Escape - Pause

Once you are done, please access the Results folder present within the main folder you downloaded, and collect the "to_deliver.txt" file. These results will help us make conclusions in our experiment, and we kindly remind you that such data will be kept private and anonymous as possible. Please, do not tamper with the contents of this file.

Results (to_deliver.txt file): *

Add file

Overall, how would you grade the tutorial you got, on a scale of 1 (Very Poor) to 10 (Very Good)? *

	1	two	3	4	5	6	7	8	9	10	
Very Poor	○	○	○	○	○	○	○	○	○	○	Very Good

Workload

In this section you will be required to assess the perceived workload required to complete Nick of Time, as well as your overall performance in the face of the tasks displayed. These will entail a variety of different aspects, such as the effort required to complete the tasks presented, how demanding it was, how frustrated it made you feel and so on. It will use a scale of 1 (very low) to 10 (very high). Above any other section of this survey, we ask that you answer as truthfully as possible to these.

Workload is something hard to measure in detail, even if simple to understand generally. As such, there are not defined rules to estimate it so to speak. You may have a very different workload experience from someone else, as there are a number of defining factors involved that sometimes are outside of one's control. It is entirely normal to feel like certain sections of experiments regarding such topics to be harder and demotivating than most.

Remind yourself that these questions are entirely subjective in this case, and must be answered from your point of view. There is no right or wrong answer, and not every single one of them might be relevant for the specific experiment at hand, and as such there is no shame in using any value present in these scales.

We thank you for your understanding.

Mental Demand *

How mentally demanding was the task?

	1	2	3	4	5	6	7	8	9	10	
Very Low	○	○	○	○	○	○	○	○	○	○	Very High

Physical Demand *

How physically demanding was the task?

	1	2	3	4	5	6	7	8	9	10	
Very Low	○	○	○	○	○	○	○	○	○	○	Very High

Temporal Demand *

How hurried or rushed was the pace of the task?

	1	2	3	4	5	6	7	8	9	10	
Very Low	○	○	○	○	○	○	○	○	○	○	Very High

Performance *

How successful were you in accomplishing what you were asked to do?

	1	2	3	4	5	6	7	8	9	10	
Failure	○	○	○	○	○	○	○	○	○	○	Perfect

Effort *

How hard did you have to work to accomplish your level of performance?

	1	2	3	4	5	6	7	8	9	10	
Very Low	○	○	○	○	○	○	○	○	○	○	Very High

Frustration *

How insecure, discouraged, irritated, stressed, and/or annoyed are you?

	1	2	3	4	5	6	7	8	9	10	
Very Low	○	○	○	○	○	○	○	○	○	○	Very High

Thank you for your participation.

Your collaboration is of the upmost importance to us, as it will make it possible to derive conclusions based on the data collected. If you have any further questions or comments, you may leave them below or by contacting us through the following email: nickoftime.thesis@gmail.com

Comments:

Your answer