PM

A Virtual Agent to Sustaining Children's Engagement in Language Learning

MASTER PROJECT

Jhair Chisuikafue Amaya Abreu

MASTER IN INFORMATICS ENGINEERING

UNIVERSIDADE da MADEIRA

A Nossa Universidade

www.uma.pt

February | 2016

A Virtual Agent to Sustaining Children’s Engagement in Language Learning

MASTER PROJECT

Jhair Chisuikafue Amaya Abreu
MASTER IN INFORMATICS ENGINEERING

SUPERVISOR
Evangelos Karapanos

CO-SUPERVISOR
Olga Lyra

Abstract

In this thesis we aimed to explore the potential of gamification - defined as “the use of game elements in non-game contexts” [30] - in increasing children's (aged 5 to 6) engagement with the task. This is mainly due to the fact that our world is living a technological era, and videogames are an example of this engagement by being able to maintain children’s (and adults) engagement for hours straight.

For the purpose of limiting complexity, we only addressed the feedback element by introducing it with an anthropomorphic virtual agent (human-like aspect), because research shows that virtual agents (VA’s) can influence behavioural change [17], or even induce emotions on humans both through the use of feedback provided and their facial expressions, which can interpreted in the same way as of humans’ [2]. By pairing the VA with the gamification concept, we wanted to 1) create a VA that is likely to be well-received by children (appearance and behaviour), and 2) have the immediate feedback that games have, so we can give children an assessment of their actions in real-time, as opposed to waiting for feedback from someone (traditional teaching), and with this give students more chances to succeed [32, 43].

Our final system consisted on a virtual environment, where children formed words that corresponded to a given image. In order to measure the impact that the VA had on engagement, the system was developed in two versions: one version of the system was limited to provide a simple feedback environment, where the VA provided feedback, by responding with simple phrases (i.e. “correct” or “incorrect”); for the second version, the VA had a more complex approach where it tried to encourage children to complete the word – a motivational feedback - even when they weren’t succeeding.

Lastly we conducted a field study with two groups of children, where one group tested the version with the simple feedback, and the other group tested the ‘motivational’ version of the system. We used a quantitative approach to analyze the collected data that measured the engagement, based on the number of tasks (words) completed and time spent with system. The results of the evaluation showed that the use of motivational feedback may carry a positive effect on engaging children.

Keywords

Children

Language Learning

Engagement

Virtual Agent

Word Game

Resumo

Com esta tese visamos explorar o potencial do conceito de "gamificação" que é definido como "o uso de elementos dos jogos em contextos que não são jogos" [30] – em aumentar concentração com uma determinada tarefa. Isto porque atualmente estamos numa era tecnológica e os videojogos, por exemplo, são uma atividade que mantêm a concentração das crianças (e adultos) por horas a fio.

De modo a limitar a complexidade, apenas focamo-nos no elemento de feedback, introduzindo-o através de um agente virtual (AV), porque o trabalho realizado nesta área indica que os AV podem ter influência no comportamento dos humanos [17], ou até provocar certas emoções através do feedback e das expressões faciais que o acompanham, que por sua vez são percebidos de igual forma aos humanos [2]. Ao combinar o AV com gamificação, queríamos 1) criar um AV que seja bem recebido pelas crianças (aparência e comportamento) e 2) possuir o feedback imediato que os jogos possuem, de modo a avaliar o desempenho das crianças em tempo real, ao invés de esta esperar pelo feedback do professor (método tradicional de ensino) e com isto fornecer aos alunos mais hipóteses de sucesso [32,43].

O nosso protótipo consistiu num ambiente virtual, onde as crianças formavam palavras correspondentes a uma imagem apresentada. Para medir o impacto que o AV possuiu na concentração, o sistema foi desenvolvido em duas versões: uma versão limitava-se a fornecer feedback de uma forma simples ("correto" ou "incorreto"); na segunda versão, o AV comportava-se de uma maneira mais complexa, onde este tentava encorajar as crianças a completar a palavra (feedback motivacional) mesmo quando as crianças falhavam múltiplas vezes.

Por último realizamos uma experiência com dois grupos de crianças que testaram as duas versões do sistema. Para analisar os dados obtidos utilizamos uma análise quantitativa para medir a concentração através do número de palavras formadas e o tempo de utilização do sistema. Os resultados finais mostram que o uso de feedback motivacional podem acarretar um efeito positivo em aumentar a concentração das crianças.

Palavras-Chave

Crianças

Aprender a Ler

Aprender a Escrever

Agente Virtual

Jogo de Palavras

Acknowledgements

First of all I would like to thank, my advisor Evangelos Karapanos that made the completion of this project to become possible, by guiding me in the right direction and providing assistance when he thought I needed it. Also to my co-advisor Olga Lyra that despite not being present for a good part (personal reasons), gave me a head start on children related subjects.

I would also like to express many thanks to Maria Ferreira, whose help in the beginning and during the realization of the thesis was crucial, and her fellow researchers at M-ITI (Madeira Interactive Technologies Institute) that helped a lot with their knowledge and hands-on, in the creation of the physical prototype, that sadly wasn't a part of the final prototype.

I must express my very profound gratitude to my family especially my mom and my dad, which made all the efforts and supported me to finish this step on my life. I am very, very grateful. Also my sister, who was impeccable supporting me and providing me with a lot of ideas.

Finally, I thanks to my colleagues for the help provided and their constructive criticism during the process of researching and writing this thesis.

Table of Contents

List of Figures

List of Tables

Acronyms

VA - Virtual Agent

VC – Virtual Character

TUI – Tangible User Interface

UI – User Interface

CSV – Comma-Separated Values

Chapter 1

Introduction

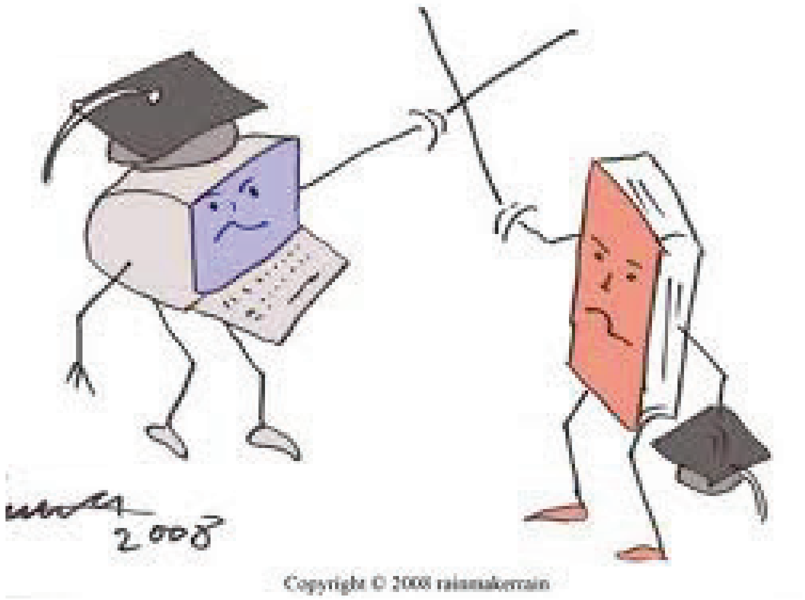


Figure 1 – Illustration on Traditional and Digital learning methods

Sustaining children's engagement with language learning is an ever-present challenge, and so there is a need to make learning activities, engaging enough to support learning. Technology has come a long way, since the first appearance of televisions to the appearance of computers. Nowadays children can spend up to 8 hours a day using different types of technology (smartphones, tablets), which is more than the time they spend at school [80].

For that reason, we need to address technology in ways that support language learning while maintaining children's engagement to the task. We can make this, by targeting the use of gamification – which is basically the use of game-specific characteristics in non-game related topics. It is a technique that has just recently started to be explored but has been proven to be successful, which we will review further in this thesis.

1.1. Problem Statement

Traditional learning methods are becoming more and more insufficient to deal with the increasing technological world we live in [54] mainly because traditional learning methods are based on learning through observation, i.e. through passive methods, and children, more than others, aren't passive learners [43]. As a result there are high levels of disengagement, which has a direct impact of the learning process. If we take a look at the possible definitions of engagement, whether as "the process by which two (or more) participants establish, maintain and end their perceived connection during interactions they jointly undertake" or as "the value that a participant in an interaction attributes to the goal of being together with the other participant(s) and continuing the interaction" [36], we can see that a very important connection exists between teachers and children, that creates learning. Roussou (2004) states that "Educational software design has attempted to include many of the tricks that characterize game design, such as the goal-directed nature of most games, ability to personalize the experience, advancement of complexity over time, etc., but has failed to equal the appeal and excitement that computer games bring to children. Hence, the division between tools for learning, represented by instructional or educational software, and tools for fun, represented by computer games, still holds."[38]. This shows that computer games have gathered all the necessary conditions to engage people for hours a day [34] and this is why we need to investigate the elements that make games what they are, and draw them out for educational purposes, so in a way to need to mimic the enjoyable and fun factors from games, and make learning more enjoyable.

1.2. Objectives

In this project we try to investigate the concept of gamification and understand how the elements that constitute it, make games attractive, but more specifically, to understand how to transmit feedback to users (children in our case). In order to do this, we need to first understand the way that feedback is provided by teachers, in a traditional classroom teaching method and in a way try to replicate it in our environment, into a Virtual Agent (VA), where we could simulate the teacher's behaviour. We decided to base our study having in mind the definition of a basic feedback experience, referred to by Cotton (1988), as a simple confirmation from part of the teacher, whether the student is right or wrong. It is important to notice that this manner of conveying feedback is better than no feedback at all. This is what we intend to replicate in our system – a feedback, but that is provided immediately (gamified), instead of waiting for the teacher to answer (non-gamified/traditional way).

The study will consist in two parts: The first is a simple feedback with no emphasis whatsoever on creating an ideal environment that motivates children to improve or keep the good work. In other words, the system works by providing feedback in its most basic way - a simple confirmation if the letters placement are right or wrong.

The second part consists on a more complex approach that does the exact opposite. When/If the child places a letter that is wrong/incorrect, our virtual character will have a more positive reaction by praising or give motivational messages such as "Yes, you got it" or "Don't give up".

Both of these approaches make use of both verbal and non-verbal feedback (facial expressions), thus making the manipulation of the facial characteristics a very important issue so that it allows for a believable appearance.

With this in mind, our goal is to understand the current teaching methods for language learning, and try to adapt them in a more technological way, so that we can achieve engagement levels that can at least be equal or better than current ones when using a traditional teaching approach.

1.3. Structure

This thesis is structured into five chapters, starting from an extensive review on existing literature, to the development and conduction of the study.

In the first chapter we contextualize our problem and describe what our goals are with this thesis.

In Chapter 2 we review the existing literature on the different aspects of this thesis, ever since the importance that literacy has in the learning process, during early childhood, to the concepts that are behind engagement - the various types of motivation and the impact they have on the phenomenon that we call Flow. We also review some gamification concepts and provide some examples of how gamification can be used and the results that these carry onto engagement. Lastly we review the current feedback strategies used by teachers and how we can implement them into our virtual agent. At the end of this chapter we provide insight on some of the existing and related works.

In Chapter 3 and Chapter 4 we describe the process behind the design and development of the system and describe how pilot tests influenced the final version of the system. These pilot tests were intended to fully understand how children reacted to the system being developed, so that we could improve it for the final prototype. After achieving the final prototype we describe the process that allowed us to conduct the field study with a population of children.

In Chapter 5 we withdraw our conclusions from this thesis, what we learned, and what could have been done. We also mention the obstacles that were faced during the development process of the system.

Chapter 2

Literature review

2.1. Literacy in early childhood

Literacy starts from the first months of life. Children already begin to learn how to read and write, way before entering kindergarten and it's no surprise that every child before going to kindergarten has already some literacy knowledge – different for everyone (some have knowledge about the alphabet; others about writing; and some even have a combination of "skills") - but nevertheless, they have some [62].

Literacy is a lot of the times influenced by the school attendance in early years and absence from school in those years can define the future of children, either school related or in life (since it is connected to learning) [71, 72, 74, 75]. For instance, researchers at Attendance Works – an organization that studies attendance effects on schools – states that one in ten first-grade students are chronically absent, i.e. miss more than 10% of school days, which puts them in risk in their subsequent years, in terms of being at a level of that current grade [76] (see Figure 2).

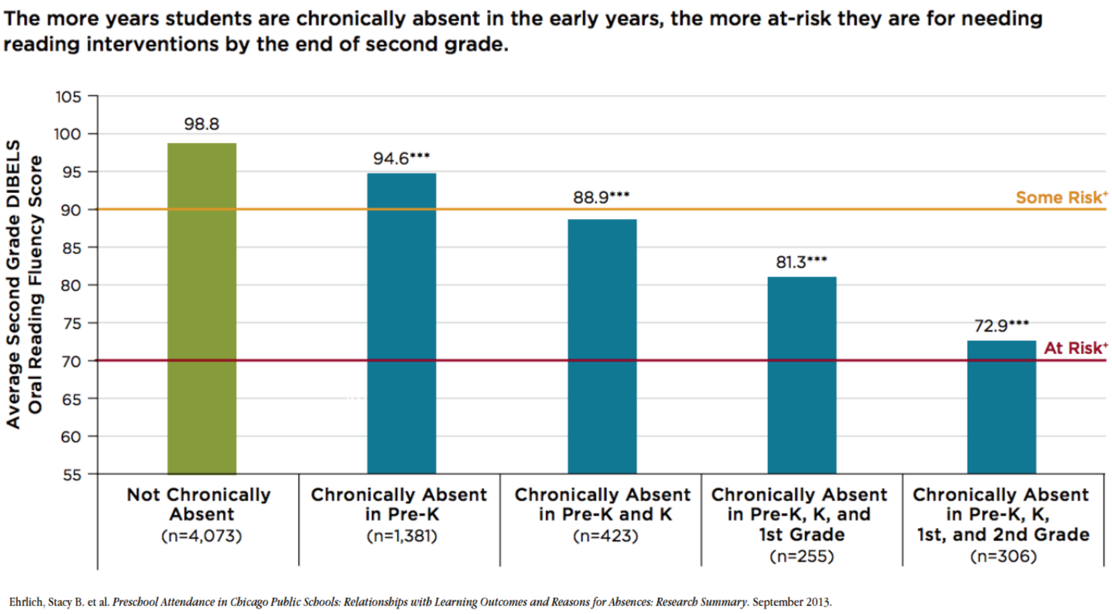


Figure 2 - Comparison of Reading fluency scores on 2nd grade, between children that were and weren't chronically-absent on previous years [76]

Magnuson et al (2006) states that one to two years prior starting school, children who are presented with learning activities, are shown to be more successful in all subsequent school years [73] in their reading and math skills. This success is even more visible in those children who come from lower economic families, since they usually have less vocabulary knowledge when compared to higher socioeconomic families [65]. More recently, Alexiu & Sodre (2011) argue for the implementation of educational interventions in the context of language learning, where we can turn cultural diversity from being perceived as a barrier, into a reinforcing lever for educational success. Therefore literacy is key, for example, on immigrant children, because of the barriers that they are faced with - linguistic or cultural. Since they aren't indigenous it's hard for them to learn new languages without proper tools. Numerous reports reveal the challenges that children with migration background face in their educational and later professional development. For instance, a recent study found that 27% of second-generation migrant students in EU countries do not reach Level 2 in reading proficiency, while this is true only for 17% of native students (Lelkes et al., 2012). In Germany, the educational performance of children with migration background has proven to be lower than their native peers and is estimated to have an impact on their educational and professional development (Frick et al 2001), and so, the use of technology can support these students in absence of a formal education.

However our focus is not only to be on immigrant's children, but to every children in general that have only basic reading skills (kindergarten and 1st grade, with 5 and 6 years old respectively), by making use of the technology that we currently have at our disposal. Labbo et al (1999) argues that technology is seen as a potential equalizer that allows all students to break their limits, which a lot of the times, are set by their poverty levels, ethnicity or gender. Technology came to change the way that children (and not only) are taught, since in the current method for learning, what they learn is already predetermined by someone, which makes them to learn at a determined schedule and at a determined time [66]. With the interactive-learning paradigm children can learn at their own pace, and by the order (of specific topics) that it's necessary, in order for them to complete a specific given task [66].

Before the mid-1980's literacy was a rarely used term, and even rarer if we put technology in the same sentence. However, nowadays with computers, this has changed. With the appearance of computers and their capability of word processing, educators interested in literacy saw potential in it as a writing tool [61] and according to Labbo et al (2007) computer-based activities can support children's literacy development by making use of several resources, such as linguistics (words; syntax), visuals (images; graphics), audio (sound effects; music), gestures (video; animations), among others [65].

A study conducted by Min (1996), aimed at evaluating the interaction between children and a learning platform, suggests that the use of technology (interactive systems) brings improvements to the child's learning ability. Min's results show that, first, age wasn't a decisive factor, because all of them were up to the task, even for those who couldn't read, and second, it was noticed that this system slightly increased children's attention span (engagement) and made them learn some new words in just one session that lasted approximately 30min. However the role of the teacher is not to be put aside, since they have a fundamental role in guiding children through their actions and answering their questions, which is essential for young learners [69], and for maintaining engagement. The problem that rises is that if, for instance, we replace the teacher by a Virtual Agent, we have no one to supervise children, and if there is no one to supervise, how can we maintain children's engagement?

2.2. Engaging children in learning

Newmann (1992) states that “engagement stands for active involvement, commitment, and concentrated attention” [79] and that “(…) we may experience varying levels of engagement as we talk, listen, observe, read, reflect and use our bodies” [79]. He also states that engagement differs depending on specific activities (for example work versus school). However, since our project is within the scope of a scholar environment and the use of technologies, the definition given by Obrien & Toms (2008) which follows as the “quality of user experience characterized by attributes of challenge, positive affect, endurability, aesthetic and sensory appeal, attention, feedback, variety/novelty, interactivity, and perceived user control” [78] suits our purpose better.

In order to obtain this engagement, we need to clarify two terms for motivation: intrinsic and extrinsic motivation. These types of motivation depend on the goal on which said motivation emerged in the first place (or the reasons that made it emerge). For instance, we could do a task because we like to do it, or we could just do that same task, but for a specific goal, such as pleasing someone (teacher for example). In greater detail, intrinsic motivation reflects a self-driven activity where there are no external interferences to the activity itself. Only something that is desired to be done by us (something we enjoy doing), without expecting any external rewards; Extrinsic motivation is the exact opposite. It refers to people that are driven by external stimuli, such as rewards (grades or even chocolates), where activities are only done for the sake of it, or competing with others (for example in games) [34, 55, 56].

“(…)From birth onward, humans, in their healthiest states, are active, inquisitive, curious, and playful creatures, displaying a ubiquitous readiness to learn and explore, and they do not require extraneous incentives to do so.(…)” [56]

Deci (1971) concluded in a study that when positive verbal reinforcement feedback was given, the intrinsic motivation seemed to increase, which is why we want to target intrinsic motivation as the key for increased engagement levels. Only then we can obtain higher quality results from the learning process [56].

After managing to increase the children's engagement, we need to know how to maintain it for longer periods of time. This brings us to a phenomenon called Flow. Flow refers to a mental state of a person when he/she is immersed in a world of their own, while performing a certain activity, meaning that their attention is entirely focused on accomplishing their current task [58].

Csikszentmihalyi's [59] model of flow (Figure 3) shows the relation between the perceived action capabilities (skills) and the perceived action opportunities (challenges).

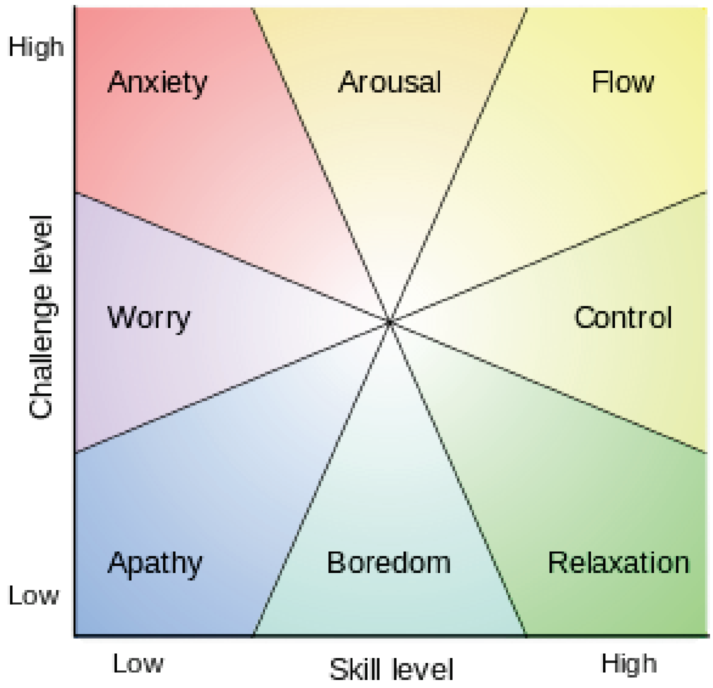


Figure 3 - Representation of Csikszentmihalyi model of flow (adapted from [59])

In this model, when the perceived challenges are higher than perceived skills, the user can enter a state of anxiety or become worried; or if challenge is lower than the skills, the user can get bored or in a state of apathy. In order to enter an optimal state of flow the challenge and skill levels must be high, i.e. the challenges by the activity being executed must be suited to one's capacities (not too hard nor too easy). This is the first condition for entering flow. The second condition is immediate feedback, where the user can understand exactly the progress that is being made, so it can adjust his actions accordingly [59].

When in state of flow, we experience the quality that better defines this phenomenon – concentration. Higher concentration makes users experience a distortion of time [59], many times ignoring basic necessities of humans, like food, water or even sleep, which is why engagement can be maintained for long periods of time.

Obrien (2008) refers a study conducted by Said (2004) in which they targeted the use of game elements in a task. For this task Said controlled the amount of interaction a child had with the system, where some would play with specific roles; others were able to manipulate the virtual word; and others were only able to watch the course of the game without intervention. Their results showed that the use of these game elements - in our case, immediate feedback - had a direct impact on engagement. So, in order to make an activity that feels like an intrinsic motivation provider, we need to take look at video games and make use of their best features, by using the concept of gamification. In the next section we explain in greater detail what this concept is and how it relates to our goal.

2.3. The Role of Gamification

By definition gamification is “the use of game elements in non-game contexts” [30]. These game elements can vary from the most simple such as score points (one of the most basic and widely used) to some more complex elements such as self-representation with an avatar; reputations, ranks and levels; teams; time pressure; decision-making abilities; feedback or even the challenge level associated to the task [39].

However, if we take each of these ‘ingredients’ and analyze them, we can see that not one can be seen as ‘gameful’, and most certainly not game-exclusive only. We cannot use a very strict manner in which we use only unique/specific game elements, simply because it would be quite constrained. In the same way we cannot use a very liberal approach (using any element that can be found in any game). So the concept of gamification tries to identify the ‘middle points’ of these elements, i.e. elements that are found in most games that are associated with games, and proven to play a major role in gameplay [39].

Throughout the years, gamification has been using game-elements (points, rewards, etc.) in order to find what increases motivation in users, for different tasks. What has been found is that experiences made, lead to more successful tasks that focus mostly on rule-based and goal-oriented play [41]. But how, we might ask? There are 3 major areas that are targeted by gamification [27]:

- 1) Cognitive - where rules define the gameplay and the challenges that arise. For example in the game “angry birds”, one can identify birds, built structures and concepts such as the laws of physics. A ‘feeling’ of desire arises in each player, to try and beat each game level. When used for educational purposes, it can transform students perception of learning, in the sense that it gives immediate rewards by completing a task towards their goal, like finishing school [27];

- 2) Emotional – where all sorts of emotions occur while playing games. Players can go from the happiest emotional state to frustration states very quickly, and in order to learn, failure must be present, until they are successful. We can relate this to education in general, where students can't afford to keep failing to succeed. Here is where gamification comes in handy - it provides some kind of resilience in case of failure, by shortening the feedback cycles that are received by rewarding the effort and not the mastery of something. In other words gamification can make “students see failure as an opportunity, instead of becoming helpless, fearful or overwhelmed” [27];

- 3) Social - In games players can try new roles. For example shy people can become leaders’ in-game, something that in real world is probably not an option for them. For students that simply don't like or can “do school” using gamification in a controlled environment can allow for students to have a social-status that can be used for social credibility and recognition for academic achievements. In simpler terms, it makes use of the needs that people have, such as “freedom of choice”, the need to be competitive with others (score boards for example), and the need to be challenged. These 3 aspects make users enter in an optimal state of mind – engagement [27].

Gamification as we can see, by bringing the features that causes games to be enjoyable, presents students with many advantages – motivating students or even provide teachers with new ways of teaching and rewards. But every good thing has its flaws. In this case, by rewarding students for each correct task, it might make them do something only when expecting something in return [27, 34]. This is an example of extrinsic motivation, as described earlier.

The study conducted by Merchant (2009) is built with a combination of all these concepts that were described so far. It consisted in a virtual world (Barnsborough) that was used by primary school students. In this world primary school students were put into play with their own avatars, and told to explore the world. By doing this they would try to collaborate between each other and construct their own narrative around multiple clues located in the world:

- Environmental clues - graffiti, logos;
- Tooltips clues - 'someone has been here';
- hyperlinked texts - 'oil drilling document' was in fact a word document;
- Interactive chat - for communication between them.

Navigational and communicational tools were built into the active world browser (Figure 4), allowing participants to walk around in virtual spaces like streets, buildings, cafes, shops and parks, and engage through written communication between them. The world was designed having in mind the places that are a better representation of the children's real world and something where they wished to live in.

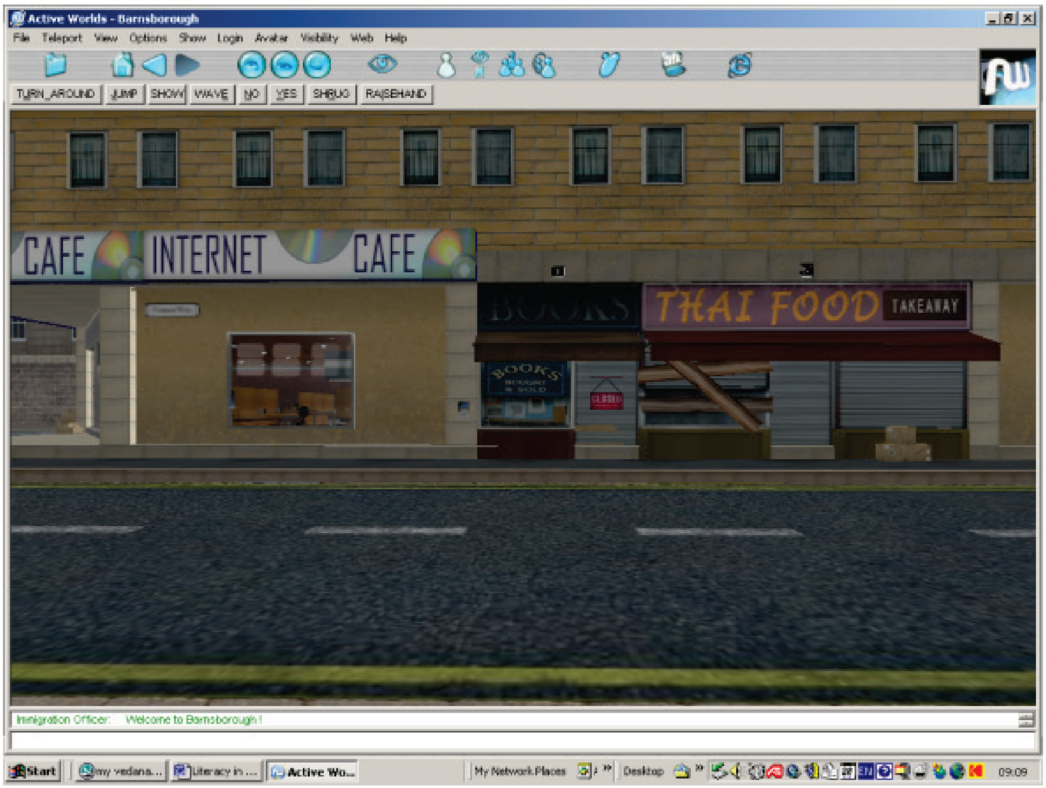


Figure 4 – 'Barnsborough' virtual world

After a while an interview was conducted with the children, in order to check on their progress. Below we can see their opinion about Barnsborough:

- "JB: its mint i like barnsbrough because its really adventurous its abosulutly brilliand MINTUS!!
- KC: its a mystery
- JM: i like it because u get to exsplor the town
- guy: um
- JB: *brilliant
- guy: what's your favourite place?
- JB: thinternet cafe´
- guy: tell me why?
- JB: *internet cafe´
- KC: town hall

- JB: i like internet cafe because lots of uknown thing have happened in there” [18]

Not only they are enjoying it, but also are being capable of auto-correcting their grammar, for example ‘*brilliant’ opposed to ‘brilliand’ and the asterisk mark. Also when asked about the chat function, they described it as “cool” and “it helps to learn how to type faster”.

The teacher, still in the virtual world, took them to the park, and told them to explore the area, but always keeping them focused with questions about the environment, to keep them ‘on task’. Below is another dialogue from one of these interactions:

- “T: . . . . remember what the focus of this lesson is! What was the park like
- before whatever happened happened?
- SS: we think it must have been busy
- T: What makes you think that S?
- CM: mm-there are some cake on the band stand so people must have been eating
- while playing
- T: Excellent observation C.
- LF: go to the poned jm
- DC: I bet some elders would admire the flowers
- SS: because it says there is a public meeting in the park
- T: What did you want me to look at/
- JB: why are other people names on avater
- T: J – you should not be messing with the avatar function. Keep focused on
- what we are supposed to be doing!

- T: There are a lot of you up in the playground. What have you discovered?
- LS: i bet the children liked the park
- CR: i bet the people would of liked the smell of the picnic
- T: Why do you think that L?
- SS: i bet people must have sat and watched the bands play and clap for them”[18]

By interviewing children, it was found that children were aware that they were doing ‘homework’/studying and that they had accepted the teaching method from the teacher, without any objections.

However, a learning activity doesn’t have to be a fully working game. Some argue that a simple score points system, ‘gamifies’ a certain activity, while others argue that it doesn’t, but the fact is, in these early stages of gamification activities, many add this ‘feature’ as a gamification element, and it does work, even if it’s for simple task. A precise case of this, is a study that was carried on children from ages 5-7 which were asked to try 2 different types of applications. One was to draw a specific gesture (letter, numbers symbols and shapes) in a blank screen of a touchscreen device for up to 20 gestures. The other application was a “target application”, where a square would appear on the screen and the children clicked on it in order to move for the next screen. Results showed that children often got bored, and there were even times in which they asked to stop or skip that specific part. Only 2 out of 7 completed the gesture task without quitting and 4 out of 7 the target task. The author then decided to add a score point system to see what kind of improvements it would bring. The score would now be updated when the gesture was drawn and the target clicked on, like many games. This score would allow in the end to exchange for small physical rewards. Straight away it was noticed that the study had improved, since the number of completed tasks went to 6 (only 1 participant didn’t finish) compared to only 2 that had completed it previously, and the only one that didn’t finish did attempt all tasks.

These studies are an example of how gamified activities manage to engage children, and thus making it a very important concept to use not only in ours, but in all educational systems. We intend to combine the strategies used by teachers in a typical classroom environment (described next in section 2.4.2) with these gamification concepts and replicate some of them (strategies) into our VA, through the use of an immediate feedback. For instance the VA in our system is capable of providing motivation by making use of the previously mentioned expressions (for example, a smile) and we do try to make a slight use of gaze, either to point something at the screen or by looking straight forward to the child.

2.3.1. 'Real-world' Projects

In the United States there is a school 'Quest to learn'[1] that teaches its students by doing game activities, like quests as their assignments, re-enact historic events and others, while receiving points for each activity.

"World of Warcraft in school" is also a project running that schools can adopt, and is based on the well-known game "*World of Warcraft"*, as a teaching method. This project allow students to have fun by learning in different areas. For instance, for practicing writing skills they have the students create a quest from scratch; for math skills students calculate average damage for different weapons and so on.

Even though they are not yet in a large scale, they are a small proof that these methods work and should be considered.

[1] Can be accessed at http://www.q2l.org/

2.4. Play based learning and feedback

2.4.1. Play based learning

Play is a term that is not yet very well defined but can be understood as a playful activity that gives us a feeling of fulfilment and self-motivation even when there are no clear goals [40, 41, 42]. It helps children to develop positive dispositions for learning, like finding an interest, the know-how and where to ask for help, being inventive (creating problems and solutions), being flexible (testing and refining solutions), making choices and decisions or even managing themselves and others [43]. It is perceived to be one of the most vital things for education from birth to 6 years old [40], making children who engage in playful experiences, more likely to have well-developed brain skills, such as memory, language and the ability to regulate their behaviour [42]. Therefore playing "engages children's bodies, minds and emotions [43]".

The best way to design a system that aims to improve engagement and learning outcomes is by performing activities that are either child-initiated or focused learning [43] (where adults guide children through the necessary tasks)(see Figure 5).

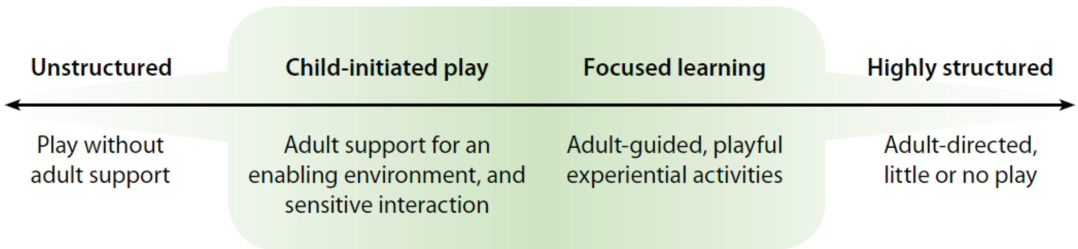


Figure 5 – Ideal learning environment for children [43]

In one end of the spectrum, we have play without any kind of adult support, and learning here is a bit limited, because even though children can play without adult supervision, it can become problematic, since children basically go in a state of 'hands-on, brains-off'; At the other end, if children perform an activity that is entirely imposed by adults, then they are not playing, which in turn limits learning. A practical example based in our system could be for instance, when a child isn't performing well in his/her task; typically when this happens, the child's engagement levels will begin to slowly go down (left-range of Figure 5), however by providing feedback when necessary (the adult being the VA), it might be possible to increase their motivation again, making them engaged.

These play activities help children develop intrinsic motivation for learning. For example they can be more willing to explore and try things out, know how and where to seek help, being inventive (by creating and solving problems), making decisions/choices or even understanding the perspectives and emotions of others [43]. Research responsible for studying how our brains are wired and how they function, shows that children are highly motivated beings that are constantly seeking for interactions - from gazing someone to asking something like 'will you play with me?'. They have in their genes the tendencies to explore and investigate the environment around them to develop their skills, and this allows them to "connect" their thoughts and their learning [43]. This confirms what Deci & Ryan (2000) said before, that we humans, are curious and want to learn. However, we should take into account a few considerations: for once we tend to use passive methods to teach children, but they aren't passive learners [43]. Children enjoy having sort of hands-on and brain-on activities, by making choices and manifesting their interests.

There have been some studies that show the importance of the whole concept of (role) playing, one of which consisted of children with ages 5-6 playing a game called "Virtual Puppet Theater". In that game they could choose 3 roles to play with: Director, Actor and Audience. What this study wanted to show was how, by playing, children could learn (in this case having different roles). For instance, by playing with each role, they started to understand the influence that their decisions have on the other roles, whether how emotional states can change or modify behaviour, and how physical and verbal actions in social interaction can induce emotions in others [21].

The point here is, the elements of play are intrinsic to engagement [78], and we want to address it by bringing the focus of our project to the 'child-initiated play' area (see Figure 5), by making them (children) run through the activity while having an adult to guide them. For this to happen we need to understand how we can create a playful activity that works as an environment provider of intrinsic motivation. We can understand it by exploring the strategies used for providing feedback, by teachers/educators in their current teaching methods and replicate some of those strategies into our VA.

2.4.2. Feedback in traditional teaching

Most strategies used by teachers were shaped by Skinner's 'Operant Conditioning' theory. In this theory, learning occurs through a process on which learning itself is associated by the consequences of a certain behaviour, i.e. the consequences caused by a specific behaviour will determine the likeability of such behaviour to be repeated or not, in the future. These behaviours can be 'manipulated' by using appropriate responses for the original behaviour – reinforcement and punishment. Such behaviours can be delivered in a positive or negative way: a positive reinforcement could be something as a reward for certain behaviours, versus a negative reinforcement which can be understood as the removal of a behaviour that has as an 'unpleasant' consequence. As for punishments, a positive punishment (usually referred to as punishment only) is a response given to attempt to decrease a given behaviour, versus a negative punishment that aims at removing the object or the something that gave rise to that behaviour (removing a children's toy for example) [77].

In the traditional teaching methods, feedback as we know it, can be delivered either verbally - oral or written - and non-verbally – which can be split into several sub-categories: proxemics – linked to the distance between the teacher and students; kinesics - connected to the way the feedback is conveyed through facial expressions; haptics – linked to the use of touch among people; among others as stated by Pan (2014). Our project will focus on kinesics because it relates for body movement, which includes gaze and facial expressions - such as a smile [45].

A typical learning environment composed of teachers in a classroom, has different categories in which feedback can be given. These can be either Rewarding, Punishing, Approving or Disapproving [47]:

- **Rewarding**: consists of mostly positive feedback. It is used when a teacher wants to reward children for their efforts in their work. It is also a part of extrinsic motivation since we can make children enjoy more the task by bringing stickers, stamps, smiley faces. However, in an experiment made with children for this type of feedback it was noticed that the use of these rewards as a promise to motivate children raised some issues. It began to be difficult to be fair to everyone and it created some sort of competition that was drifting away from the goal of the experiment [47].

- **Punishing:** as the name indicates, it's mostly about negative feedback. Teachers here show their disapproval for something children did. For the same experiment used in the rewarding type, here teachers either moved children to other place or send them out of the classroom. This resulted in a notion of being cast out of the classroom community and support from part of the teacher [47].

- **Approving:** related to when a child's work exceeds the teachers expectations, which causes them (teacher) to give an approval feedback due to their work or engagement. It can lead to rewards that are present in rewarding feedback, but is often considered rewarding itself, the fact that teacher approved their work, and for that reason is associated with positive feelings for both of them. The most important factor here is the nonverbal communication that exists when giving this type of feedback [47].

- **Disapproving:** very similar to the punishing type - children are mostly the ones to blame for something they did or didn't, like concentrate, or try to complete their work. In the experiment this type of feedback was also used in front of the whole class, as a way to transmit authority and try to reinforce socially acceptable behaviour. It was never used when the teacher saw that the child at least tried. Few characteristics of teachers, to transmit disapproval include eyebrows raised or lowered; the tone and volume of the voice; pointing to something without saying anything [47].

Despite some of the types of feedback being negative, they are crucial for supporting learning. In some cases negative feedback can be positive, as shown by Tunstall & Gsipps (1996) [47]. They claim that giving individual feedback to a child in a public way (in front of the whole class) carries dangerous consequences, however, in certain cases by doing so, it would take the attention to them (children) to carry on a discussion around something, which in turn created an opportunity for learning.

Cotton (1988) states that the most widely used technique that teachers use to deliver feedback is praise, however, it was found that it's not always beneficial, and depending on the situation, can be helpful, neutral or detrimental [50]. This praise by the teacher doesn't necessarily reinforce learning, but it has a positive impact on students, either by filling the emotional needs or managing their (children) behaviour [50]. He cites some strategies, from Slavin's (1986) book "Educational Psychology: Theory into practice", that teachers can use in classes and that work as positive reinforcers. For instance teachers use things like a wink, a smile, an "I'm proud of you" or even writing "good job" in their tests. More recently Pan (2014) conducted an experiment that explored the way teachers taught their classes and also came up with a few similar strategies that allow for improving the interaction between teachers and students. A few examples were to get students motivated with a smile, making them feel happy and confident; use of eye-contact because it established the sense of presence while reinforcing the importance of the message the teacher was trying to transmit; be happy for children's interest in something; when teaching language, draw attention for specific letters and words and give opportunity to students to talk about what is read and the sounds of them; engage in play-based activities, such as writing words [48].

These strategies are the ones that we aim at exploring, since they are the ones that we can identify as being closer to our system, and therefore have a system that is a more accurate representation of the reality.

2.5. Play and gamification

By combining all that that has just been said, the whole idea of the concept of play is to let children drive their own activities and not imposed entirely to them by us, adults. They have to experiment and keep trying new things, communicate their ideas or feelings and learn from their own mistakes [43]. A very effective way for them to learn is precisely by them taking risks and learning even when they fail, which will eventually allow them to acquire new experiences where learning will naturally occur. Usually children who respond/do something incorrectly, tend to attribute their failure to their abilities (or lack of) rather than blaming the teacher and that, in order for them to increase their learning, they should always be active, and not just listening, mainly because “memory is the residue of thought” [52] and so in order for them to remember they need to think.

2.6. Virtual Agents and gamification through verbal and non-verbal (kinesics) feedback

Up until now we have explained the concepts of gamification and play-based learning, as a way to increase learning outcomes and how these are related to intrinsic motivation. We have also seen that ideally what we want to create is a system that is supported by the theory of a child-initiated activity. But the question here is how can we develop a system that can motivate children to complete a certain task? First we have gamification that is responsible for keeping child’s engagement, but then, there is also the intrinsic motivation that can be provided by ‘playing’. With this is mind we intend to implement the VA in a way where it can act as a replacement of an adult (something that technology is aiming for [81]), and help children when they need. First we start by reviewing a few general characteristics of virtual agents, and see what kind of impact they have on humans, and then we talk about how we can make them provide feedback in a way that sustains engagement.

2.6.1. Anthropomorphic virtual agent

Anthropomorphic virtual agents (‘Human-like’), as proven by numerous studies, have had a big impact on increasing engagement. They are a powerful ‘tool’ that can be designed as social models, for specific situations. For instance we can design them to positively influence young children to stay in school, by making them (Virtual Agents) to look like someone that children can look at - a ‘role model’. The VA’s are not only introduced to virtual environments to make them appear more realistic. Lots of the time they are introduced in order to create content and mood [22] and according to Nass’s paradigm - where computers can be interpreted as social actors - there has been evidence that humans can be influenced by these agents, just as they would by human social models [17]. The appearance is perhaps the most important factor. VA's have been found to have equal levels of influence as humans in inducing certain emotions on the counterpart, and its facial expressions are interpreted in the same way as those of humans, while attaining the complexity and diversity of those of humans [2]. An example of this is a study that showed when users saw an avatar that looked like themselves exercising and losing weight, they started to exercise more in the real world [17]. This shows us that by making good use of them, we can design VA's that can influence humans' engagement.

Figure 6 – Examples of Anthropomorphic Virtual Agents (from [17])

2.6.2. Virtual Agent and Feedback

There is considerable research that says that feedback leads to learning gains [53]. Following the points that we made in the above section, we can see that the way they (virtual agents) communicate with the user is very important. Vinayagamoorthy, et al (2006) state that nonverbal communication can override verbal communication, i.e. a negative message (conveyed in a nonverbal way), that is given with a smile, can be perceived as friendly, even while not being so. In essence, nonverbal communication is taken as the true psychological state of an individual, especially when the cues are negative [22], and in order to create a believable social character, nonverbal communication must be part of it, even being a complex task. For instance, looking at someone in the eyes can be characterizes by either loving or aggressive, but it all depends on the relation between both [22]. These expressions represent how we, humans, truly feel about something, since we can say what we'd like, but our face won't look same if we don't mean what we say.

There is also the behaviour associated with the VA, which can be more important than the visual realism of the character, since people expect the characters to behave according to their appearance, and any discrepancies will be disturbing [14] (because people often respond to praise, criticism and social responses from a computer the same way they would respond to a human. [17, 22]). A similar study to ours, conducted by Prendinger et al (2005), tried to analyze the impact that affective-feedback and non-affective feedback had on stress and frustration levels of users. Basically they had a VA that had a certain level of empathy associated, that was used in the respective versions of the system (affective feedback and no-affective feedback). For example, in affective-feedback version, the VA had two emotions associated for each type of answer (correct, incorrect and even a 'delayed' answer); when correct answer was given by the user, the VA would show appropriate facial expressions for the correct answer (happy). Similarly would happen for an incorrect answer (sadness). They concluded that the empathetic version of the character against the non-empathetic version, significantly lowered the perception of difficulty and frustration levels, by just representing a correct body language (smile or appear sad depending on answers) and by providing some kind of feedback when there was a delay associated with an answer.

Another study conducted by Slater, Pertaub & Steed (1999), consisted on asking people to talk in front of a virtual audience and then were asked how they rated their performance. Their responses were clearly affected by how the virtual audience gave positive or negative feedback. Even sometimes, when the audience was negative the perceived audience interest overcame the negativity. The positive audience caused greater self-rating and reduced the speaking anxiety. Both these studies show the impact that a VA has on humans, and for those reasons, we need to combine all the reviewed methods used by teachers, in order to provide this feedback in the best possible way.

2.7. Related Work

2.7.1. GraphoGame

GraphoGame [34] is a study, made also for children, that has a different approach but similar goals to our study. It makes use of sounds of letters to make children learn which letter that sound corresponds to. By hearing the sound, they have to choose between some letters available in the screen.

Figure 7 – 'GraphoGame' UI

In order to evaluate the results of the study they set up a class with small groups of people where each of those classes had the duration of 45min, 4 times a week, during 28 weeks (7 months). This allowed them to see how this technological way of teaching/learning compares with the traditional methods.

The goal of GraphoGame was to assess how certain features can retain children's engagement, specifically the rewards system, and the challenge the game brings (difficulty level).

Results of that study show that children that had a reward system initially started by having high interest but as sessions went on, it reduced to almost same level than children that didn't had any rewards system, until eventually differences were null. According to the parents, children stopped using the game because they eventually learnt how to read, and therefore there was no point in continuing.

In a version of the game there was no feedback about the progress or skill development, which in turn did not help children understand why they were playing the game.

As for results of this study, the difficulty level didn't had any impact on the engagement and as we could expect the children that used GraphoGame represented the one that had more gains.

Similarly in [35] the same game was used for evaluating neural processing in children. For the period of 8 weeks that GraphoGame was used it started to used specific parts of the brain that is responsible for reading and in about 3 weeks children started to improve their phonological processing and writing of words and pseudo words.

2.7.2. The virtual quiz master

In contrast with 'GraphoGame' this work had a different approach and slightly different goals to ours. The 'virtual quiz master' [23] had the goal of measuring not engagement, but the frustration levels of users, when presented with a specific situation. However the main focus of it, is that in order to measure the differences between frustration levels, it relied on the feedback of a virtual agent.

The study consisted on a series of mathematical questions, where users would sum 5 displayed numbers, and then asked to subtract a number (i, where i<5) of that exact sequence. But in some versions, the 5th number had a delay on showing, which caused some frustration (users wanted to be the first to finish and with best scores).

Results showed that if the VA showed empathy (apologizing for the delay) and presenting correspondent emotions (for example bowing down to say sorry and a sad facial expressions) versus just replying ‘correct’ or ‘incorrect’ it had significant effect on lowering frustration levels.

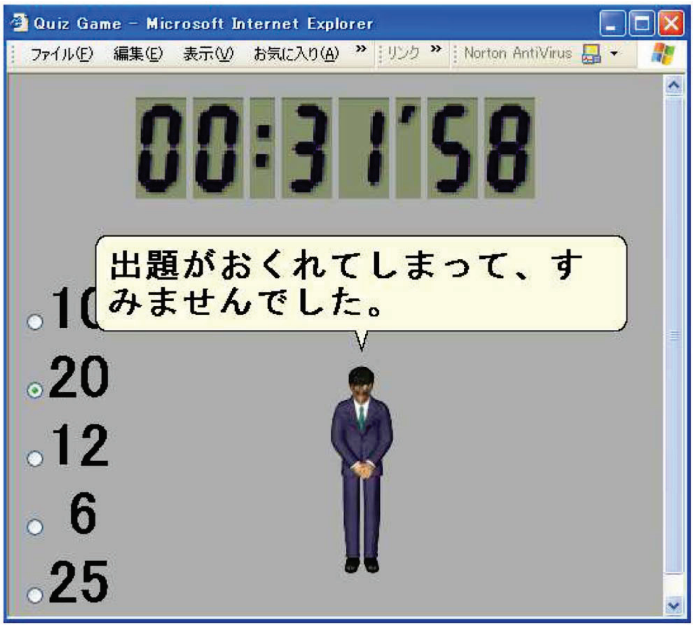


Figure 8 – The ‘virtual quiz master’ apologizes for delay

2.7.3. Agents Appearance

Baylor et al (2005) manipulated various characteristics of a VA, more specifically ‘attractiveness’, ‘coolness’ and age in order to investigate some student’s motivation, regarding the possibility of choosing an engineering career. They were then given 16 possible agents to choose from, and it was noticed that women tended to choose the agent that was female, attractive, ‘cool’ and young as, in their words, being ‘more like themselves’. Men in the other hand chose a male, older and ‘uncool’ agent. After receiving persuasive messages for a period of time they found that the ones who had the possibility of choosing an agent, were more influenced than others. Base on this study we can infer that people can be more influenced by someone who is alike to themselves. But we should bear in mind that this is not always the case. The same author states another study that had as goal verify the change of behaviour that the avatar could have. For that they manipulated the attractiveness of the avatar, and it was noticed that participants with good looking avatars, demonstrated more friendliness and extroversion. A similar thing happened in an experiment conducted by Bailenson et al (2001) that consisted interacting with a virtual environment through a head mount, and measure the distance on which they walked around an avatar. This avatar, in one version, was a self-representation of the user, while in the other version (control) that avatar was someone unknown. What they concluded was that users would walk much more closely to their virtual-selves (M=0.23m, SD=0.15), when compared to virtual-others (M=0.51, SD=0.19). More recently, Rosenberg-kima et al (2007) concluded from their experiment that when an agent was visible, students tended to have greater levels of motivational outcomes.

These studies demonstrate the power that an avatar can have in a virtual environment, and that appearance has a major role on humans. It is very important that virtual agents are designed having in mind their ultimate goal, i.e. if we want to impact a certain group of users, for example – motivate young ones to stay in school – we should design one agent with all the features that are found to have influence for such group of users. In this case agents could be as an older and 'cool' person, while being careful in all the details, such as clothing, gender, ethnicity and even their voice.

2.7.4. SAM

Sam [1] it's a conversational agent whose appearance is destined for pre-school (~6 years old) children. It was designed with advanced speech strategies, meaning it had a more rich vocabulary. Its goal was to see if a virtual agent with richer vocabulary (quoted speech – "Oh no, she said"; temporal expressions – "today is …"; or spatial expressions – "from the *other* side") could impact the learning outcomes, by collaboratively telling stories. The system was able to track the child's movement and listen to sounds. When someone arrived near SAM, she would salute them and tell a story of their own in a hope to provoke a first engagement with the child. When the story was done, SAM would then place a figurine (which acts like a 'token', i.e. whoever has it, tells a story) in a castle compartment (a physical structure represented in front of the child). As soon as it was the child's time to tell a story, he/she would open the compartment of the castle, where it would find the figurine, and then tell a story, where SAM was constantly listening and saying things like "what happened next?".

Figure 9 – Visual representation of SAM

As a result of this study, it was noticed that children, despite being engaged, were using more of the special and temporal expressions, as well as the use of quoted speech, as the study went on. These findings show that virtual agents can enhance learning, hence it contributed for increased literacy levels.

2.7.5. Animated agents in interactive learning environments

Almost two decades ago, Johnson and Rickel (1999) analyzed some studies that made use of virtual agents as a way to promote learning, which by then, were very few. The most complex one that they found was called "Herman the bug" and had as a purpose to learn about the field of botanical anatomy, where Herman (the virtual agent) monitored students' actions and gave according feedback, by explaining concepts and hints. In that system, middle school students were given a set of environment conditions, where they had to 'create' a plant that derived from other plants and could survive in the proposed conditions. The system was split in various versions where one gave only visual advice, other gave verbal advice only, and a third where we had a combination of both. Upon executing pre and post-tests to evaluate learning eventual learning outcomes, they concluded that students that used the mixed approach of the VA, had significantly better results in their post-tests, which suggested that a well-designed agent that can make use of verbal and non-verbal actions, can bring beneficial results for learning.

As we can see, and based on the previous sections, we can confidently say that the authors' assumptions were indeed correct, and is explained with the success that this approach has had for learning environments, in the last decade.

Chapter 3

The design and development of CLEVA

With the development of CLEVA (Children's-Learning Engagement using Virtual Agents) our goal was to create an agent that was able to relate with the aspects that we found to be of particular relevance on existing research both on virtual agents and gamification, and thus increase children's engagement in the best way possible. Upon analyzing and reflection of the literature, we knew we needed to have a friendly avatar (VA) that behaved accordingly to its appearance, so that children wouldn't be disturbed by their actions. We also needed to create a pleasant environment both visually and non-visually (audio) in order to make children feel good. To do this we set a few goals that needed to be accomplished, in a mission to create a VA that reunited the best aspects for increasing engagement. With this said, our goals were to 1) create an agent (VA) with an appearance suited to our target population (children); 2) create an agent that behaved accordingly to the situation (context-dependent); 3) create a system that supported a child-initiated play approach; 4) create an agent that was able to convey verbal and non-verbal feedback in the best way; and 5) a 'pleasant to the eye' environment.

With these goals in mind, this section provides a detailed description of CLEVA, ever since its first iteration up until the final one, along with why and how certain decisions were taken.

For starters, the creation of the whole system was possible by using *Unity* software. The virtual agent was only possible by using a software that allowed for manipulation of the facial expressions. Such software was *Crazytalk7* by *Reallusion*.

The base for the final system design was provided by the pilot study that was made before conducting the final study. These aspects/decisions are described thoroughly in the next section 3.2, but we can summarize these aspects on alterations revolving around the number of animations that existed, the replacement of some of the objects that the system provided, for example, substituting a dog for a more pleasantly looking dog (see Figure 10), and making some tweaks in our code to increase the idle timer, so the system became less interruptive.

Figure 10 – Evolution of the images shown to children

After all these changes were done we obtained a system with a general workflow as represented in Figure 11, through a flowchart.

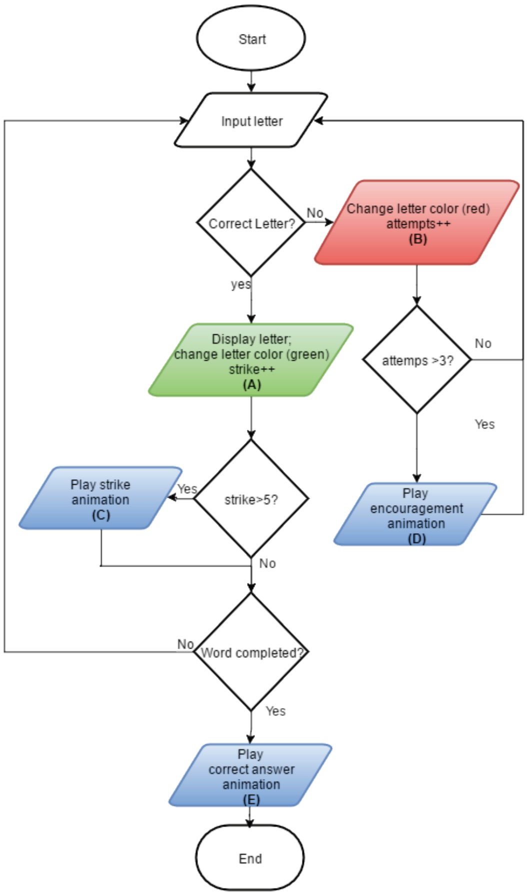


Figure 11 – System's Flowchart

Figure 12 – Visual representation of flowchart. Top-left image corresponds to (A); Top-right corresponds to (B); Bottom-left correspond to screenshots of (C) and (E) animations; Bottom-right corresponds to (D) animation

As we can see in the flowchart, our system is continually monitoring the child's activity and their answers throughout their interaction with the system and based on these parameters we are able to infer when to intervene, so we could provide the immediate feedback that we talked about initially on the gamification concept. In this case (for the motivational version) as soon as children fails for three consecutive times, an animation for encouragement is immediately played, and the same happens for a strike animation – in this case five consecutive correct letters. This feedback acts as reinforcement to make children more motivated and continuing the task with more enthusiasm.

3.1.1. Virtual Agent and Facial Expressions

The character chosen to be the VA of our system was the one illustrated in Figure 13 below. This decision was made because there is research that claims that virtual characters can become more believable if the child manages to 'identify' with the character, hence we chose a young and friendly-looking avatar [15]. With this said its appearance takes into account our initial goals, where 1) it has a friendly appearance, 2) it's suited to our age range (young) and 3) it had an age-appropriate voice (for verbal communication). For the verbal communication we had a child with the age around of our target ages to pre-record some sentences that were later used with the animation.

Figure 13 – Virtual Agent

In order to make the virtual character more believable, we have to target the facial expressions along with the feedback provided. But we might wonder, why do facial expressions matter if we can provide feedback verbally or even written? As we have addressed before the non-verbal communication (i.e. facial expressions) usually outweighs the verbal communication [14, 22], and to verify this we can address the description given above for the workflow of the system. In that specific case if a strike situation is present, by simply giving feedback with no expression at all, it can be a bit disturbing – which goes against the goals set for the system – because verbally the VA can show energy and motivation but if that isn't accompanied with proper facial expressions it can be an odd thing to see. That is why facial expressions are important and are a must when delivering feedback.

However, there isn't currently an actual consent on which expressions are the most basic/universal, but there is a wide consent by researchers that all possible expressions are derived by a set of expressions that are divided between fear, disgust, sadness, anger, enjoyment, surprise, happiness, [2,3,4,5,7,14,16,22].

With all this in mind the need to focus on the facial expressions was high, and so it was decided that we needed to start from the basic facial expressions mentioned above. Nevertheless, it's important to notice that for the scope of this project, where 'positivity' is something that we aimed at, the less positive expressions, such as fear, disgust or even sadness are rarely used, despite making some appearances in the control version of the system.

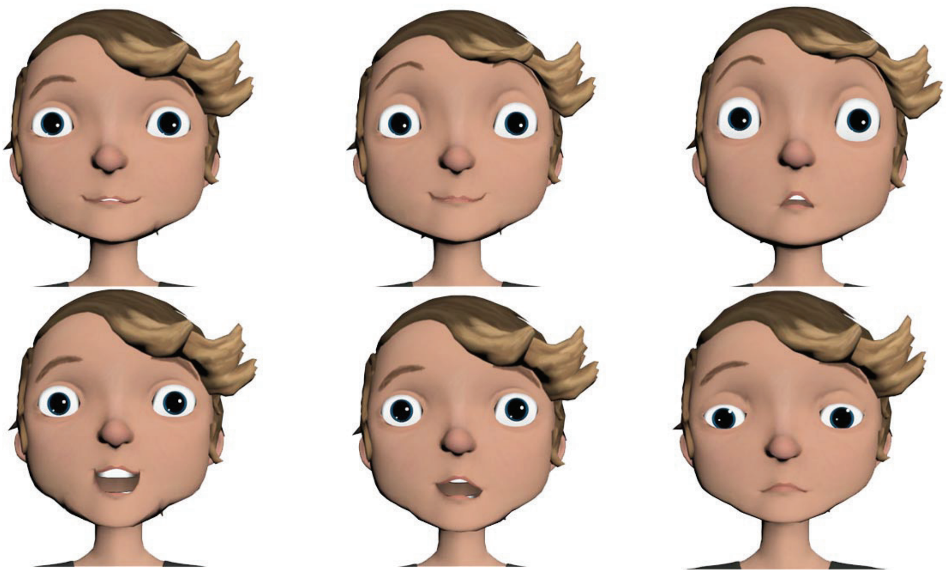


Figure 14 - Facial expressions with some variations for 'happy' (left and middle columns), surprised (top-right) and sad (bottom-right)

In Figure 14 we can see the facial expressions that our VA is capable of conveying and are a result of an analysis on the characteristics that are paired with their respective emotions. For example, a surprised expression is characterized by an open mouth, eyelids and eyebrows pulled up (see Figure 14 top-right image), or for a sad expression where the corner of the lips are pulled down and the inner part of the eyebrows are slightly pulled up. These expressions were created by making use of adequate software - CrazyTalk7 – whose customization for each facial aspect was vastly available. It allowed us to move the needed facial points in the character, to achieve a 'model' for the facial expression in the character. In Figure 15 we can see some illustrative examples of the facial expressions configuration and the crazytalk7 customizability (Figure 16).

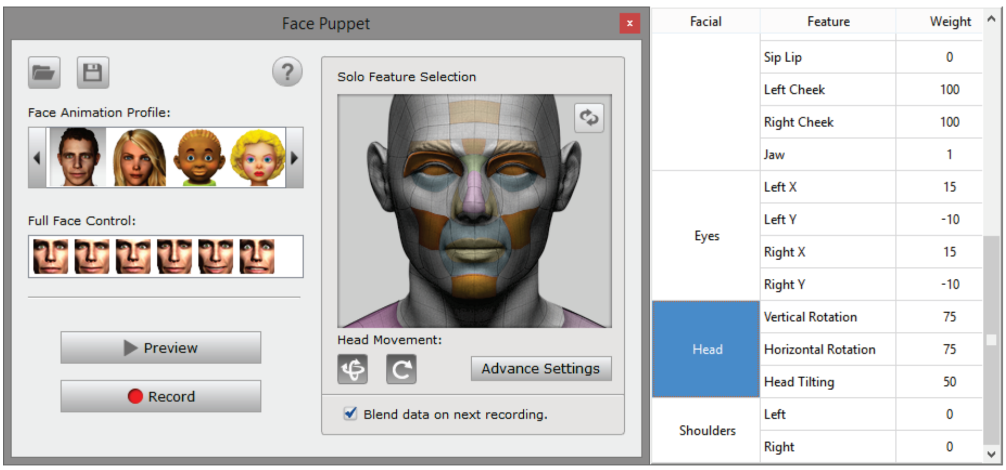


Figure 15 – Manipulation of general facial areas and weights for each area

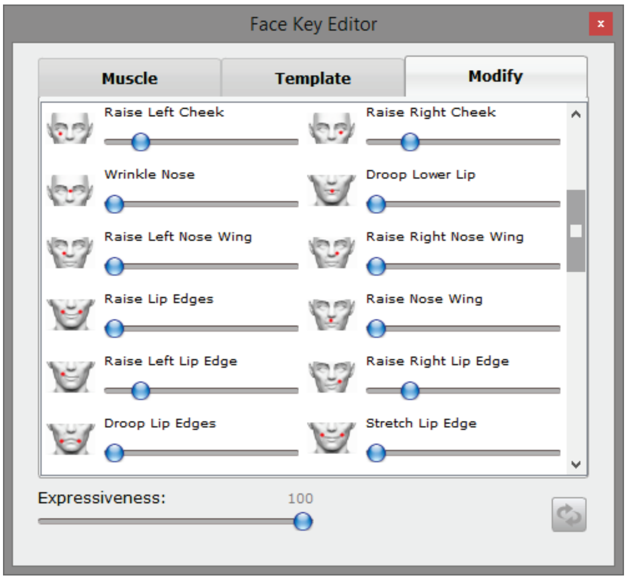


Figure 16 – Manipulation of more specific facial parts

3.1.2. System Design

The system is constituted by a home screen, and the screen where all the experiment consists on. The home screen is only intended for the evaluator as they choose which version (defined previously) of the system will be played in each group of children.

In order to obtain a graphical interface that was minimally child-oriented, it was necessary to recur to image editing software such as Adobe Photoshop and Adobe Illustrator, mainly because the images found on the internet that were best suited had a watermark, that made the background unpleasant. Figure 17 shows the final result for this home screen.

Figure 17 – System Home Screen

For the main screen, where all the interaction happens, it was necessary to recur to the same tools as for the home screen for the images and additional software for the manipulation of audio and video animations. Upon loading, children were presented with a welcome message, where the virtual agent presents itself, talks briefly about the goals of the game, and tells what they need to do in order to advance. This is a very important step as it's the first time the child sees the VA, and a good first impression is crucial for them, whether it is to create a 'connection' so they can relate themselves with the agent physically or simply to like the agent. In the tables below (Table 1 and Table 2) we can see the animations that the virtual agent is able to provide for each version of the system for all specific situations:

Table 1 – Animations executed by the VA for the motivational version of the system

Type of Animation	Animation Content
Encouragement (multiple incorrect letters in a row)	Come on, you can do it! Don't give up! You were almost there, try again!
Strike (multiple correct letters in a row)	Yay you're doing good! Hooray! You got another right! Another correct letter! The word is almost right! You're doing very good!
Correct Answer (correctly formed word)	You're the best! Amazing, you're right! Perfect! Wow, excelent! Iupi! Very good! Yay you got it! Awesome!

Table 2 - Animations executed by the VA for the non-motivational version of the system

Type of Animation	Animation Content
Correct Answer	Right Correct Ok, it's correct
Incorrect Strike (multiple incorrect letters in a row)	Wrong That's not ok Wrong letter Eh, it's wrong
Strike (multiple correct letters in a row)	N/A

As for the game, the objects/images are presented in a wooden sign next to a chalkboard, where the word being formed appears. These objects, as a result of the pilot study (discussed next in Chapter 4) were carefully chosen to be friendly looking and easily recognizable for children of the targeted age.

The decision for the final layout (as seen in Figure 18), was made considering the fact that the virtual agent is composed mainly by its face, and since it had no body attached a floating head didn't look very good if it were to stay above the letters (which in turn would be located in the bottom). All these elements (sound effects, audio, images, videos/animations) were added because Labbo et al (2007) stated that these aspects are proven to contribute to literacy improvement, especially in computer-based activities.

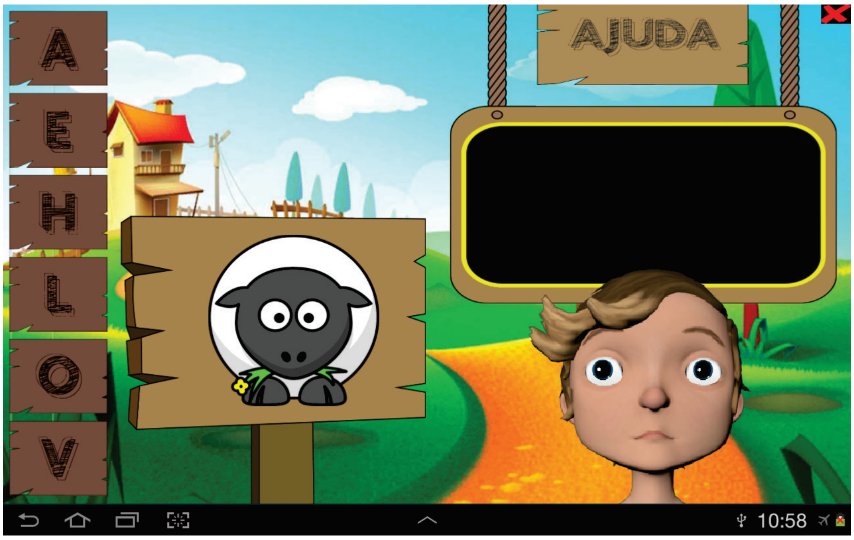


Figure 18 – System's main screen

Lastly, in order to get the data for posterior processing (statistically-wise), it was implemented an exit button that was placed in the top-right corner. When this button was clicked information about the interaction process (see Section 4.3 for the exact logged information) between the child and the system, were carefully saved in a text file following the CSV format (Comma-Separated Values), so we could have a way to easily export the data to Microsoft Excel, where later on we could analyze and handle the necessary information for the results representation.

The final 'product', as seen in Figure 18, was a result of our initial goals where we defined as priorities the agent's appearance; its behaviour (proper feedback with correspondent facial expressions); a good looking environment where all the elements blend in with eachother; and a system that supports child-oriented play, by letting children learn though their own actions (visual feedback on letters - Figure 19).

Figure 19 – Button feedback – neutral (left), incorrect (center) and correct (right)

3.2. Formative evaluation of CLEVA

In this section we present the pilot study that was executed in order to gather more information on how children interacted with the system and get some insights on how to improve it, by observing how their reactions were concerning the events that occurred when interacting with the system. Its obvious purpose was to understand what part of a learning system, children value the most and improve/build a prototype from there.

For this first phase, we involved a total of 4 children with ages from 4 to 7, where 3 were males and 1 was female. Their age range was part of our strategy to find the most suited age range for children to use these types of systems, and having chosen a range allowed us to see what the ideal age was. Apart from this, it was important that children knew how to function with a smart device (tablet in our case), which they all did.

Figure 20 – Conduction of the pilot study

The main purpose of this preliminary study was mainly to observe children as they interacted with the system. We used an Unstructured Passive Observation [68] in order to register their behaviour. We took notes by recording the interaction so we could analyze it later and know what points we missed. After this test was concluded we conducted a kind of mini-questionnaire/interview, where we asked 2 questions, where we could obtain information that wasn't possible by just observing their interaction:

1. "Did you understand what the boy (VA) said?"
2. "What did you dislike from it?"

The system used for this pilot study was constituted simply by the objects that the child had to form the word about, the virtual agent (with an idle animation in the background), a background where we could connect all this, and a basic set of animations.

Figure 21 – Pilot study system

Figure 22 – Pilot study system

Concerning the age ranges for the study, what we can say is that the lower and upper limit (4 and 7) were found to be the less ideal age range for our system, mainly because, the youngest one (4 years old) didn't really know much of the letters, which made him just randomly click on letters not knowing what he was doing, and therefore rendering the experiment useless on him. However it's important to notice that by the end of the experiment he could assimilate some specific letter that he couldn't in the beginning, suggesting that he at least learned something with the system, which confirms previous work on vocabulary learning, such as SAM's system (see section 2.7.4). The oldest one (7 years old) in the other hand, was found to complete the tasks very easily. Children around that age have their writing or reading skills in a typically more advanced level, and for that reason, they are more familiar with word forming, therefore showing very little difficulties. This left us with the ones that were 4 and 5 years old, that as we could verify were the ones that 'struggled' a bit more, due to their not so advanced writing skills.

As for an analysis of the two questions asked to children upon finishing the experiment we concluded a few things. For the first question answers were straight forward and all children, with no exception, understood what the VA was trying to say. This was a plus in our point of view, mainly because the addition of the 'voice feature' for the system was somewhat complicated at the beginning, and was only accomplished by managing to get a child to record some voice content for the project. With this in mind, we could confidently afford to use the audio for our final system, by making only a few minor improvements where needed;

The second question was a little more in-depth and we wanted to see if they (children) could give us some hints/guidelines on which we could work on, however upon briefly analyzing this question, we came to a conclusion that maybe it wasn't the best question to ask, perhaps because of the 'complexity' that it carried onto younger children, since the ones that didn't answer were the youngest ones. Nevertheless, the other kids answered directly saying that they liked the music (which suggests that there weren't apparent dislikes). One other child answered that it was too easy, and as we stated before, it was mainly due to his age (7 years old).

After watching their interaction and analyzing their answers we concluded a couple of things: 1) by understanding what the VA said, it showed us what we had seen on existing literature, which was the engagement that the agent managed to create with children. Some of the children even knew that the avatar was trying to help them; and 2) just as Csikszentmihalyi's flow model demonstrates, when a task is too easy (low challenge level and high skill levels) the user enters in a state of boredom or relaxation, which is precisely what we verified with the oldest child.

As for the virtual environment itself, we found the first major problem to be associated with the animations that were played. This because, due to the nature of which they were played (randomly), it would sometimes play the same animation more than once in a row, and when this happened, children would often notice and say that the character was saying the same thing again, and a wooden behaviour makes the believability factor to disappear, and it is not what we wanted, as stated by Vinayagamoorthy et al (2006).

Our second major problem was the way in which our messages where given to the test subject, i.e. when he/she chose multiple letters correctly in a sequence (what we call a 'strike' – see Table 1 for reference) we had the animation 'popping' out of nowhere, which we found to interrupt the flow of the activity, and by interrupting this flow, the optimal state for engagement – concentration - was nowhere near to happen [59]. And again, this wasn't something that we'd want, but the way the system was designed (due to constrains – explained in the Conclusions chapter) was the only way to provide the feedback to children, and we never thought it to be a negative aspect, until performing some pilot tests, where we observed some frustration related behaviour when the animation was given. This was a really big red flag that had to be worked on for our final prototype.

Another problem that was encountered was the fact that the end of the inactivity animation and the beginning of the same animation later on, was spaced to close to each other, i.e. when a child was idle for too long, it triggered a animation to be played, and this did manage to show some signs of annoyance in children.

A situation that was disliked by some children was the fact some of the objects were a bit scary for them, which implied a need for replacements of said objects in the scene.

Another thing that we noticed is that the "Exit" icon/symbol, where children could click in case they wanted to finish the experiment earlier, wasn't quite visible because it was blended in the background (bad color combinations).

Last but not least, some reactions that we got from children testing, was that they did pay attention to the present virtual agent, and when asked in the end if they understood what the character said, they replied “yes he was saying that I was wrong sometimes”, and were euphoric with some of the object representations, for example the “fish” or the “mouse”, which suggests that they were fond of their ‘cartoonish’ representation.

These problems helped us model future iterations, by indicating us that a few improvements were necessary to be done. For starters, we had to come up with a way of providing the motivational feedback in a way that it didn’t disturb the flow of the task. Also, we needed more animations in order to decrease the probability of the same animations being played again (repetitive) – which contributes to a wooden behaviour; We also needed to increase the inactivity timer, so children could have a longer response time, before playing again and therefore become less annoying. Moreover, in the design process it was noticed that the application looked very basic, i.e. it was too simple, and with that in mind some minor animations had to be implemented for future versions of the system.

Chapter 4

Summative Evaluation of CLEVA

This chapter explains in detail how the evaluation of the system is to be conducted, from the tasks necessary to test the prototype, to the roles of all of those who were involved directly or indirectly in the study and we also address the methodology used.

First, we started by taking into account the issues that were raised while pilot testing the system. For instance, our most serious problem, as mentioned in section 3.2, was related to the disturbance of flow when children were performing their task. With that in mind we decided we needed to create some alternative where flow wasn't disturbed, which in turn allowed us to get more accurate data. As a result we ended with a prototype that made use of both a computer and a tablet, where the computer displayed the necessary animations, instead of being played in the tablet screen, and thus putting an end to the 'interruptive' experience. Further in this document we'll go into more details on how it was done.

4.1. Goals and Research Hypothesis

The main goal of our evaluation is to find out if motivational feedback manages to impact engagement to the task - which translates to a longer usage time of the system, while committing less errors and consequently, know if this motivational feedback shortens the time needed for children to form the correct word. We chose these as the main aspects to measure engagement because they go into account of the engagement definition given by Obrien & Toms (2008) (see Chapter

2). With these goals in mind we postulated some hypothesis to on which we base our work on:

Hypothesis 1: The presence of gamification through motivational and immediate feedback will lead to prolonged periods of time using the system (higher engagement);

Hypothesis 2: The presence of gamification will lead to the completion of more tasks;

Hypothesis 3: The presence of gamification will lead to a faster completion of the same tasks;

Hypothesis 4: The presence of gamification will lead to fewer errors while performing a task;

Hypothesis 5: The presence of gamification by means of a virtual agent will lead to a more pleasant experience for learning.

4.2. Defining test tasks and roles for adult participants

In the context of our project, a task represents an image/object that children need to form the word about. For instance, when presented with the image of a dog the children's task is to form the word 'dog'. This corresponds to a single task. Our system is composed of 50 images, and therefore 50 tasks and so, children will use the system until it eventually reaches the end – completion of 50 words – or until they decide they don't want to continue, either from being tired or for any other reasons that they might have.

After clarifying what a task is, we then need an adult that is responsible for guiding children should they ask. For instance, it was noticed in pilot tests, that some children didn't know what a determined object was, and at that point they asked "what is this?". For this reason we added a button 'Help' which sole intention is to try to help children understand what the image is about. When this button is clicked the VA intervenes and gives a hint on what the object is. As a last resort, if the child failed to understand, then it is time for an adult to intervene and help the child with the image. This feature was added having in mind the research already done on these fields that state that both play in general and intrinsic motivation are closely related, mainly because they are the combination for engagement to happen [78], and since our system is targeting the child-initiated play field (see Figure 5) it needs to be able to let children play on their own, by allowing them to learn by their own choices and mistakes. Failing is vital for learning [27, 43, 52] and only with minimal interruption from part of the adult, is that children can develop intrinsic motivation [43].

We could have added a ‘skip word’ feature, but it is in our best interest that children form the words, and not skip them, as the purpose of the project is to keep them using the system longer, allowing them to enter a state of flow where they can eventually learn new things.

The other task that adults need to realize during the conduction of our study is mostly for the purpose of the posterior phase of data analyzing and conclusions withdrawal. The supervising adult is responsible for taking notes of children’s behaviour, such as when they demonstrate happiness or concentration.

4.3. Data Collection

The system is constantly logging the actions taken by part of children. Among the logged actions is the following data:

- Total time spent using the system – measured from start to finish;
- Exact time spent for the completion of each specific task/word;
- Average time spent to complete a task;
- Average time spent for choosing the correct or incorrect characters;
- Number of times the child was ‘idle’ (no activity for >30seconds), which triggered the inactive animation.

In addition to these metrics, and in order to get accurate measurements we need to reduce external influence to the system, like parents or someone helping. This was thought to be necessary because Markopoulos et al (2008) stated that in order to reduce the distraction factor it was a must to use a free area, with minimal distractions possible, and interestingly enough, we actually found it to be very true while pilot testing. Clearly when other children were present, the 'tester-child' would often get distracted and interrupted by them, which caused us to intervene and separate them. For this reason the study was thought to be conducted individually.

Some other popular metrics that can be used are provided by Markopoulos et al (2008), and can be divided into multiple categories [68]:

1. Observation methods – Here the interaction is observed either in a structured or in an unstructured way. The main difference is that in a structured way the evaluators observe a set of behaviours that are already established beforehand, while unstructured is observed the natural interaction that occurs with no pre-defined behaviours. This last one, is particularly good for finding unexpected issues and also usage patterns;

2. Verbalization methods – Children during the conduction of the study verbalize their thoughts so the person evaluating can know exactly why children find the system difficult (if it happens). Other way is the intervention of the evaluator, that intervenes to ask questions about specific things of the interaction process;

3. Survey methods – Constituted by interviews and questionnaires. Questionnaires has the advantage that we can use paper forms containing rating or ranking scales. Also it can save time, when compared to interviews, because it can get lots of data from multiple children. Interviews are meant to be conducted one-on-one with children and can allow for a more in-depth discussion with each child.

Of these, we came to a conclusion that the most suited to our purpose, are based on observation and survey methods, mainly because we want to know what children feel like during the test, without disturbing the flow of the task, and to accomplish this, those methods are the ones that can give us a more detailed answer. However, they claim that conducting an interview with children is more prone to incorrect data mainly due to the fact that an interview is usually carried in a one-on-one situation, and so, it can make children feel uncomfortable and provide unreliable answers (very dependent on the present people or the surrounding environment).

Figure 23 - Smileyometer

The use of graphical representations or scales are also another popular manner of retrieving the desired information. An example is the well-known smileyometer (see Figure 23). The smileyometer provides 5 options, where each option represents the rating of the experience, ranging from a sad face (awful) to a more happy face (brilliant). We intend to use this method as a way to know exactly if the children that participated enjoyed the task and therefore being more receptive to these types of learning environments.

These metrics were selected as the best ones for a better and more detailed analysis, to be performed later in this chapter for the study conduction.

4.4. Data Analysis

In order to support the veracity of the proposed hypothesis, the obtained data was analyzed having in consideration multiple factors. For starters, we needed to exclude all those who didn't meet our criteria for testing the prototype, and this was made possible through an interview with the participants, even knowing some of the disadvantages these interviews carry (as mentioned above) [68]. Nevertheless, we found this (interview) to be the better way to detect for abnormalities in literacy levels among children that could otherwise have influence on our results.

In a later stage (after collecting all necessary data) we carefully considered all details (specified in the above section) from which we calculated from a statistical point of view, the values of means, standard deviations, correlations and some independent samples T-tests. This allowed us to compare both versions and posteriorly relate them with all our proposed hypothesis and also for a better understanding of the data itself and the patterns that emerged from the obtained information.

4.5. Methodological approach

In order to get accurate results we considered only the results of children who were up to the task, i.e. children who had the minimal literacy knowledge. This decision was mainly because research [68] suggests that when dealing with children, it's important not to exclude any children in order to not make them feel bad with themselves and thus leaving them wondering why they weren't allowed to do the same task than others. Markopoulos, et al (2008) state that in order to address this problem, we should carry the study with everyone, and then exclude results from those who were not fit to the task. The information relative to those 'fit to the task' was obtained through small questions of the alphabet, i.e. the recruited children (specified in the next section) were shown the alphabet and were then asked which letter was which. Afterwards we asked them to spell a word provided by us. Those who didn't meet our criteria (knowledge of the alphabet), were 'flagged' so we could later (during the data analysis) remove their results.

During the testing, we used an Unstructured Passive Observation approach [68], where we observed the entire interaction with the system, while looking for behavior related to our goal, through note taking. This way we managed to get information that later was used to uncover the most recurrent behavioural patterns, more specifically the ones that showed signs of whether the child was focused on the task (enjoying it), or even bored (like being forced into doing it).

In order to fully analyze the usage of the system, we used a brief questionnaire constituted by a few questions carefully created having in mind Markopoulos et al (2008) guidelines because we're working with children and the questions need to be as simple as possible. As a result these questions allowed us to understand the children's points of view of the system they used:

- "Did you enjoy the game?"
- "Did you feel that you learned something new?"
- "Do you prefer a real teacher or an in-game teacher?"
- "Did you prefer not to have an avatar (Virtual Agent)?"
- "Did the avatar help you to continue forming words?"

This questionnaire (in a sheet of paper) was individually handed to each children in the end of their experiment, and made use of the smileyometer for all the questions except question 3 "Do you prefer a real teacher or an in-game teacher?". All other questions answers were measured using a Likert scale ranging from 1 (negative) to 5 (positive). Figure 24 shows the questionnaire given to children:

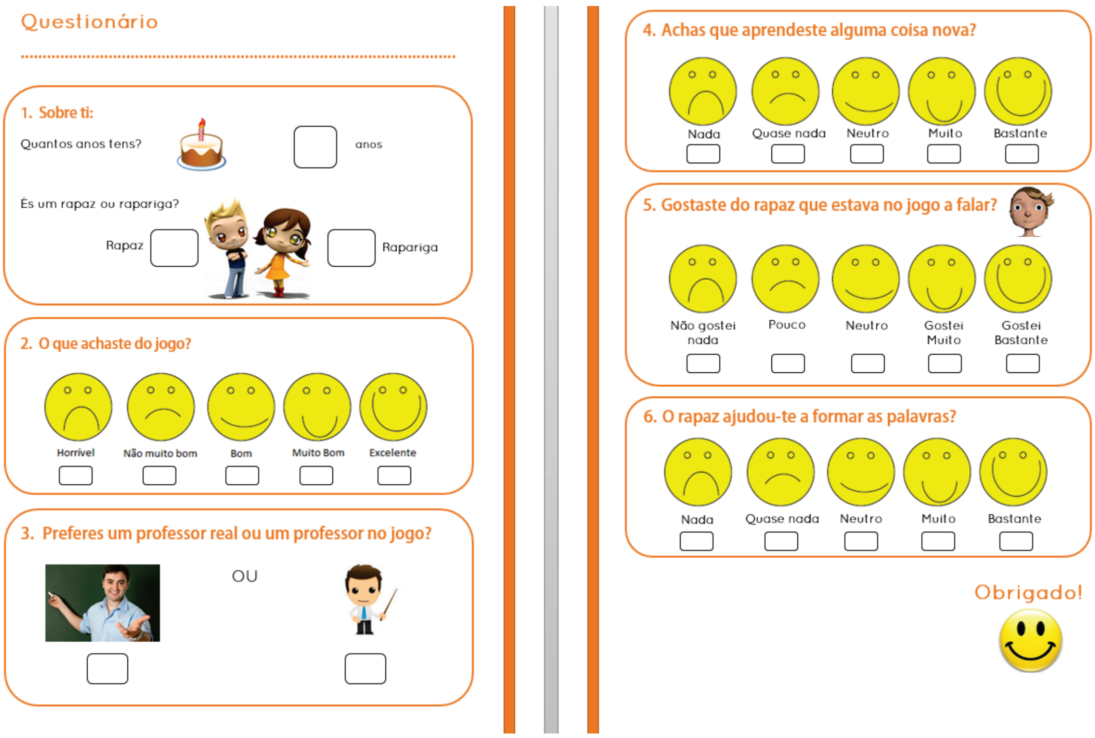


Figure 24 – Questionnaire to be filled by the participants

4.6. Participants

As mentioned previously, our target population were children whose ages were among 5 and 6 years old. Ideally the recruitment for participants would be children of a school classroom, however due to technical problems, this wasn't possible, and so we needed to look individually for children. After we found some, we reached out to their closest friends, and so on. We made sure we had authorization from the parents who were very receptive to the study.

The participants had the basic knowledge of letters (Alphabet). The chosen age range was selected having in mind that children of those ages are usually the ones that have the basic knowledge for words reading and writing. According to our pilot studies, children that were younger than 5, didn't really know the letters of the alphabet and therefore couldn't participate in the study. On the other hand, children older than 6 were already somewhat fluent in the mentioned skills (reading and writing) which would make the study 'useless' or drift apart of the main goal, hence the choice of the age range afore mentioned.

During the course of 7 days we had a total of 18 children participating in the study (12 males, 6 females), of which 8 were found to be completely knowledgeable of the alphabet and 8 were familiar with most letters of it (mostly the younger ones). Only 1 child owned a computer at home while the others didn't own one, but had access to one, and also similarly 8 owned a tablet, and therefore were familiar with the use of both of those devices. The mean age for the participants was 5.7 years old, with a standard deviation of 0.47.

According to the methodology defined in the previous section, all children were asked to participate in the study freely, meaning that those who didn't want to participate, they wouldn't be forced to do so.

4.7. Study Design

The subjects that participated in our study are categorized as 'in-between' subjects, i.e. there are multiple participants that are allocated in each version of the system.

In sum the system consists of two versions:

- The first version involves a system that provides motivational feedback (Motivational Version) to the subject, when this is thought to be necessary. In other words, if the child currently using the system keeps failing, the system intervenes (by the virtual agent) and provides motivational messages (see Chapter 3 for a list of the motivational messages) in order to avoid losing the attention of children. The global functioning of this version can be seen in Figure 11.

- The second version (Control Version) is based on the same system, however it does not make use of motivational messages, i.e. the VA still provides feedback but in a non-motivational way. Figure 25, represents the general workflow for this version.

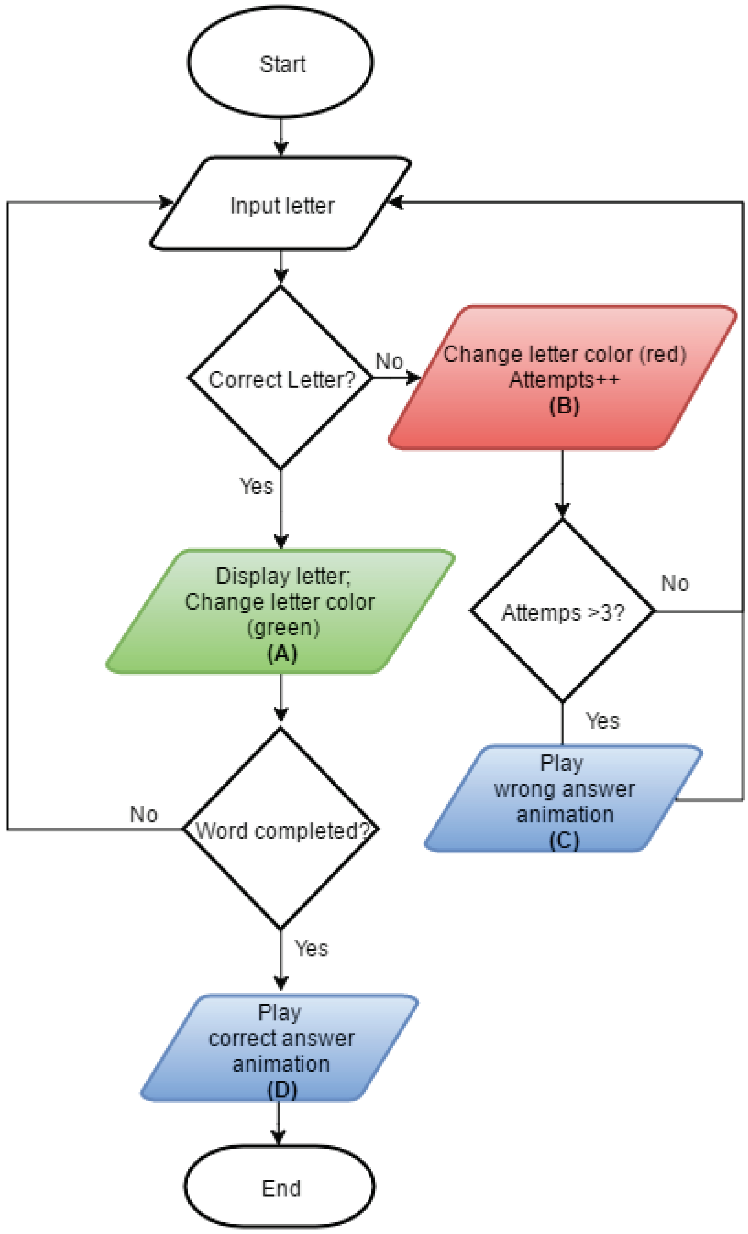


Figure 25 – Flowchart of control version of the system (non-motivational feedback)

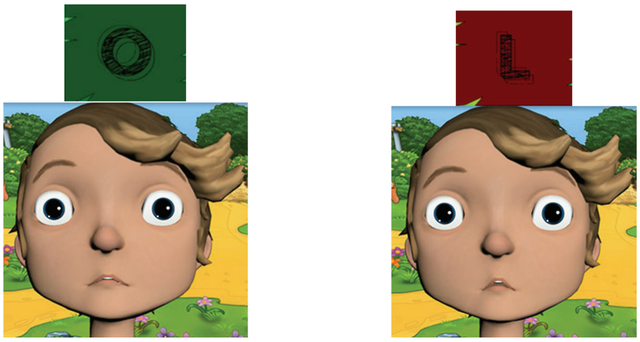


Figure 26 - Visual representation of flowchart. Top-left image corresponds to (A); Top-right corresponds to (B); Bottom-left correspond to screenshots of (D) animation; Bottom-right corresponds to (C) animation

It's relevant to say that the given feedback was a result of the investigation around how feedback is delivered in the traditional methods and the best way to deliver it. In our motivational version for example, we wanted to take into consideration the use of both positive forms of giving feedback: rewarding and approving [47]. As we explained in our literature review chapter (Chapter 2) the reward type (characterized by small rewards like stickers or smileyfaces) targets mostly extrinsic motivation, and that is exactly the opposite of what we wanted with this project which targets intrinsic motivation. So, we made use mostly of the approving type where we give positive messages to the user as a way to congratulate their achievements, while having attention to the voice tone. It is also important to notice that this type of positive feedback relies heavily on nonverbal communication [47], which further contributes to the importance of the facial expressions, as we referred in the previous chapter on the importance of facial expressions.

Also, both systems were designed having in consideration the gamification aspects defined by Lee & Hammer (2011) - cognitive, emotional and social aspects. The cognitive area is targeted through the use of some basic rules that define the gameplay, i.e. children can't choose the letters that were already correctly placed, and when these letters are chosen, changing their colors to green or red (for correct and incorrect letters, respectively). The emotional aspect is mostly used in the feedback that is provided by the VA, when it is given in a (for example) failure situation, by providing appropriate feedback which in turn could make them less afraid of failing again. The social aspect is the only one that we didn't address, because its major features were not in the scope for the goal of this project (e.g. various roles, scoreboards, competition factor, or even freedom of choice – most of which are characteristics of a full game). Some might think that scoreboards could have been added for the purpose of competition between other children, however since we wanted to analyze the impact of feedback specifically, if we added scoreboards, we could not have been able to say for sure which one had influence in the results.

Nevertheless, both these versions are fully functional interactive prototypes that are intended to test the ability that a virtual agent has on engaging children.

4.8. Study Procedure

The first and most obvious task was to find children willing to be subjects to our system. After finding them we then needed to find a location. The location chosen to conduct the study was mostly in a building where children gathered around to play. We also did conduct the study in a couple of the participant's house (due to the proximate relationship between some of them). This was mainly because of the familiarity they had with the environment and with this permitting us to get accurate information.

After having the participants and the location we needed to get the necessary material to carry the experiment. Initially we were going to use just one device for testing both versions (we only had one available), however upon reading Markopoulos et al (2008) suggestions for study conduction we realized that we were going with the bare minimum and if for some reason it failed we would find ourselves in an inconvenient situation. With this in mind we set to find more devices as a way of redundancy. We ended up with having two android tablets that were later assigned to specific locations, relatively far apart from each other, as to minimize the chance of children 'glancing' at each other. This setup permitted us to save a lot of time during the testing because 1) the use of only one device was going to be very time consuming, and 2) more than two devices would force us to have more adults to be present and would also create some unwanted confusion in the room, both with the adults and children. Each device had both versions of the system installed, providing this way redundancy of the systems in case of failure of a device. Also for redundancy purposes we had available another two tablets capable of running the same system.

Lastly, before carrying on with the study, children were asked if they knew how to operate the tablet and were then explained on how the system worked and what their role in the experiment was. After this was settled children could start to play. In the cases of having two participants at the same time, we provided one of the participants with some headphones (see the left participant on Figure 27), so they wouldn't be interrupted by each other's devices. We were careful enough to clean them (headphones) after each session, for hygienic purposes.

Figure 27 – Two participants carrying both versions of the study – Control (left) and Motivational (Right)

The experiment ended when/if a child wanted to stop playing. Following this we provided them with the questionnaire (see Figure 24) and while they were filling it, we started to prepare for the next child (if there was other one waiting). When the child finished filling the questionnaire we thanked them and their parents for participating.

Figure 28 – Study conduction

4.9. Results

After concluding the experiment we proceeded to analyze the obtained data through the system which allowed us to make a quantitative and qualitative analysis. The quantitative analysis served for understanding if children in the motivational version had performed better or worse than those who participated in the control version of the system. A qualitative analysis in the other hand permitted us to understand what the good (what children liked) and the less-good features of the system were.

Table 3 – Statistics for the time spent with system statistics

Group Statistics						
	Condition	N	Mean	Std. Deviation	Std. Error Mean
Total duration (min)	Control	9	13,0978	2,87877	,95959
	Motivational	9	15,5811	6,49985	2,16662

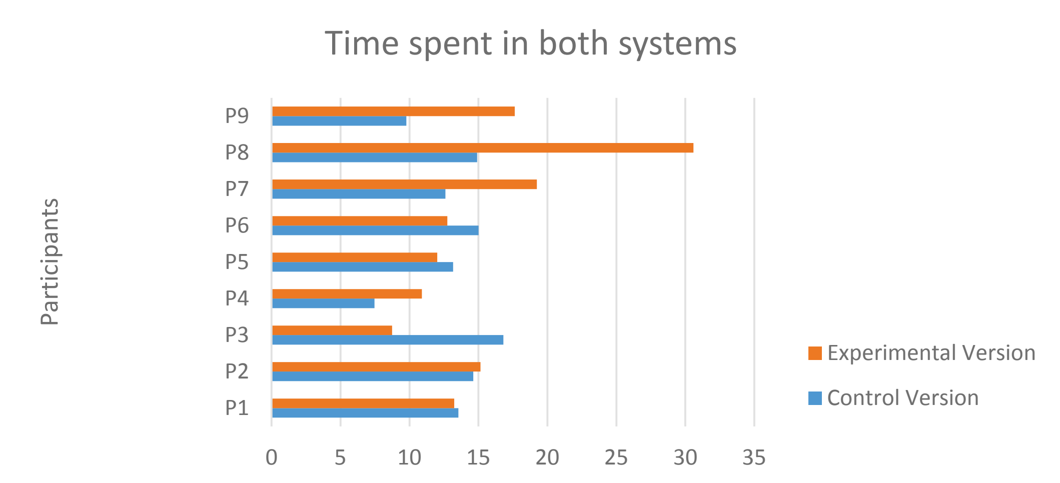
Our first impressions were that our motivational version seemed to have better results than the control version, even though we expected the difference to be more accentuated. This was first verified by looking at the total time spent (TTS) with the system. The motivational version had a mean usage time of approximately 15.58 minutes when compared to the 13.09 minutes accomplished by the control group. The results however were more consistent within the control version (13.09 minutes) versus the 15.58 minutes of the motivational version, i.e. the control version results were mostly around the mean value (SD=2.9) whereas motivational version had a more diversified data (SD=6.5). This suggested that the motivational feedback version had a slight impact on the length of the activity. However, when using a statistical approach we came to the conclusion that (statistically) the difference was almost nonexistent. We concluded this by carrying an independent samples t-test where we compared the time spent with the system with motivational and non-motivational feedback conditions (Table 4).

Table 4 – T-test results for time spent with system

Independent Samples Test										
		Equality of Variances		t-test for Equality of Means						
									Interval of the	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Total duration (min)	Equal variances assumed	2,588	,127	-1,048	16	,310	-2,48333	2,36961	-7,50667	2,54001
	Equal variances not assumed			-1,048	11,022	,317	-2,48333	2,36961	-7,69752	2,73085

We found there was not a significant difference in the scores for the control version (M=13.1, SD=2.9) and the motivational version (M=15.58, SD=6.5); $t(16) = -1.05$, $p=0.310$. This suggests that despite the difference noticed in the beginning, statistically, motivational feedback didn't influence on time spent with the system. With this said our first hypothesis (H1) - The presence of gamification through motivational and immediate feedback will lead to prolonged periods of time using the system, can be considered valid, even if statistically wise this difference is not very significant. However, based on the research that we previously analyzed, we expected this difference to be much more distinguishable, because multiple studies [18, 56, 78] showed that feedback (sometimes paired with other elements) had direct impact in engagement, which leads us to believe that this difference was due to the fact that it was about something new, where initially the curiosity and the 'novelty' factor played its role.

Figure 29 illustrates the time spent for each individual participant for each version of the system:



	P1	P2	P3	P4	P5	P6	P7	P8	P9
Experimental Version	13.24	15.15	8.74	10.9	12.01	12.74	19.23	30.59	17.63
Control Version	13.54	14.62	16.81	7.46	13.16	15.01	12.61	14.9	9.77

Figure 29 – Comparison between time spent on both versions of the systems (in minutes)

Secondly, we looked at the average number of tasks completed by each group of participants that similarly appeared to be in our favour - ~39 tasks (SD=8.84) versus ~36 tasks (SD=9.98). We noticed during the testing that the participants who did the least amount of tasks, were the ones that according to the questionnaire, enjoyed the game less than others, or who found the VA to not help them form the words, and as a consequence they spent little time with the system and completed less tasks. However we faced the same situation that happened for the previous hypothesis (H1). We conducted an independent-samples t-test where we checked the number of tasks completed with motivational and non-motivational feedback conditions (Table 5).

Table 5 – T-test results for number of tasks completed

Group Statistics						
	Condition	N	Mean	Std. Deviation	Std. Error Mean
Number of tasks	Control	9	35,89	9,981	3,327
	Motivational	9	38,89	8,838	2,946

Independent Samples Test										
		Equality of Variances		t-test for Equality of Means						
									Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Number of tasks	Equal variances assumed	,006	,941	-,675	16	,509	-3,000	4,444	-12,420	6,420
	Equal variances not assumed			-,675	15,769	,509	-3,000	4,444	-12,432	6,432

We found that there was not a significant difference in the scores for both the control (M=35.9, SD=9.98) and motivational versions (M=38.89, SD=8.84); $t(16)= -0.675$, $p=0.509$.

Nevertheless, this data shows that even though statistically the difference isn't very significant, we did obtain a higher number of tasks, which verifies our second proposed hypothesis - The presence of gamification will lead to the completion of more tasks. However, one of our assumptions in the beginning about the time spent using the system increasing the number of tasks completed was wrong, because after calculating the correlation between the TTS and the number of tasks, we ended with a rather low correlation ($r=0.16$) which tells us that there was not a significant relation between these two variables (0.1 being very low and 1 being very high) that influences the number of tasks completed.

For our hypothesis 3 - The presence of gamification will lead to a faster completion of the same tasks – and even though we 'contained' the two ends of the spectrum (lowest and highest) of literacy level presented by each individual child (as described in the methodological approach section) we noticed that some children were more skilled than others (naturally), which made them complete tasks faster. An example of this is participant nº 15 (P15) that managed to complete almost all tasks (48) in 12 minutes when compared to participant nº19 (P19) that only achieved 33 tasks in 17 minutes (see Table 6 and Table 7 below for a full reference of this data).

Table 6 - Data obtained for the control version of system

	Control version								
Participant	P1	P2	P3	P4	P5	P6	P7	P8	P9
Age	6	5	6	6	6	5	6	6	6
Sex	M	M	M	M	F	F	M	M	F
Nº of completed tasks	30	35	48	15	34	41	35	48	37
Total time spent (min)	13.54	14.62	16.81	7.46	13.16	15.01	12.61	14.9	9.77
Nº of idle (>30 secs)	4	4	0	1	3	2	3	2	0
Mean time per task (secs)	25.8	24.81	20.55	28.16	22.88	21.54	21.54	18.2	15.74

Table 7 - Data obtained for the motivational versions of the system

	Motivational version								
Participant	P11	P12	P13	P14	P15	P16	P17	P18	P19
Age	5	5	6	6	6	5	6	6	6
Sex	M	M	M	M	M	F	F	M	F
Nº of completed tasks	48	48	21	38	48	38	35	41	33
Total time spent (min)	13.24	15.15	8.74	10.9	12.01	12.74	19.23	30.59	17.63
Nº of idle (>30 secs)	0	1	0	0	0	2	3	7	2
Mean time per task (secs)	16.17	18.48	24.1	17.09	14.66	19.28	32.85	44.67	31.1

With this in mind we first needed to get some fair values for later comparing the two groups. This is mainly because not every participant completed the same number of tasks, and also since the difficulty level was a not a factor to consider in this project (i.e. longer words sometimes are harder to spell, which takes longer to answer), results would be affected and henceforth calculating an average value for the time of each would produce incorrect values. To repair this miscalculation we made a comparison between the timestamps for each individual letter/character, rather the full word. This way we would remove both the different amount of tasks completed and the difficulty factor, out of the equation, and therefore obtaining more precise results.

As a result, even by making those calculations, we surprisingly found very similar results (Table 8), i.e., the difference between the two versions was of less than a second (more specifically of about 0.4 seconds).

Table 8 – Average time needed for completing tasks in both versions

	Motivational Version	Control Version
Mean time for each letter (seconds)	3.75	3.39
SD for each letter (seconds)	0.89	0.99

With these results it's fair to acknowledge that motivational feedback hasn't had impact on how fast words are completed, which rendered our third hypothesis (H3) as invalid.

Our hypothesis 4 - The presence of gamification will lead to fewer errors while performing a task – in the other hand, has proven to be successful. For this hypothesis we initially calculated the average of the number of incorrectly clicked letters, to see which version produced a greater number of incorrect clicks. As a result we had a mean of 50.9 in the motivational version, opposed to 75.8 of the control version. In order to verify the impact that our independent variable (motivational feedback) had on the number of errors in both versions, we also conducted an independent sample t-test (Table 9).

Table 9 – T-test results for the number of errors

Group Statistics

	Condition	N	Mean	Std. Deviation	Std. Error Mean
Number of errors	Control	9	75,89	29,759	9,920
	Motivational	9	50,89	19,903	6,634
Average time for incorrect letter	Control	9	10,7656	3,37791	1,12597
	Motivational	9	19,0044	10,38501	3,46167

Independent Samples Test

		Equality of Variances		t-test for Equality of Means						
									Interval of the	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Number of errors	Equal variances assumed	,437	,518	2,095	16	,052	25,000	11,934	-,298	50,298
	Equal variances not assumed			2,095	13,963	,055	25,000	11,934	-,602	50,602
Average click time for incorrect letter	Equal variances assumed	5,444	,033	-2,263	16	,038	-8,23889	3,64019	-15,95574	-,52203
	Equal variances not assumed			-2,263	9,674	,048	-8,23889	3,64019	-16,38693	-,09085

We found that there was a significant difference in the errors made in the control (M=78.89, SD=29.76) and the motivational versions (M=50.89, SD=19.09); t(16)= 2.095, p=0.052. This shows us that children are prone to make less errors when given feedback in a motivational way.

We also found of interest to compare these errors, with the average time it took for those errors to happen and as Table 9 demonstrates, there is also a significant difference for the latter (M=10.76, SD=3.38 for the control group and M=19, SD=10.39 for the motivational group); with t(9.67)= -2.263, p=0.048. Both these results suggest that children think more before choosing a letter, rather than choosing one randomly without putting much thought, which is something very positive, as we can verify by the studied literature [43] and can be related as the 'hands-on brains-on' activities that cause intrinsic motivation.

Lastly, to verify the validity of our final hypothesis (h5), we needed to approach it using both forms of analysis, i.e., the questionnaire (quantitative) and our observations on the interaction with the system (qualitative). First, the individual questionnaire consisted on 5 items of a Likert scale where participants could evaluate their experience with the system. Each item (i.e. question) was based on a 5-points answer, as we can see in the table below (Table 10).

Table 10 – Classification for questionnaire questions in a Likert scale

Question 2	Points
Horrible	**1**
Not very good	**2**
Good	**3**
Very Good	**4**
Excelent	**5**

Question 4 and 6	Points
Nothing	**1**
Almost nothing	**2**
Partly	**3**
Helped	**4**
Helped a lot	**5**

Question 5	Points
Not at all	**1**
Not much	**2**
Partly	**3**
Liked it	**4**
Liked it a lot	**5**

Question 3 of the questionnaire is not represented because it had as a purpose to understand what type of teacher children preferred – real or virtual – and such, it didn't meet the criteria for representation on the Likert scale. However we could say that based on their answers, the majority of children prefer to learn from a 'virtual being' – 12 of them versus 6 that rather have a real teacher. This finding suggests that children enjoy their virtual experiences enough as to receive instructions from them, as was stated by existing research [17] (more on Chapter 2 for literature review). For full reference of the questionnaire answers see appendix A.

As for the other questions for a total of 18 participants (20 initially but the results from two participants were discarded for being too 'low' in literacy levels) - 9 participants in each version (n=9). Table 11 shows the obtained results.

Question 2 (Q2) results, shows that children enjoyed playing the game, regardless of the versions they participated in, being the average answers similar on both versions (control version – mean of 4.0; motivational version – mean of 4.1). The proximity of these values suggests that even though with the 'lack of functionality' of the control version children enjoyed the game as a whole and thus future applications of this kind are prone to be well received by them (children). Question 5 (Q5) results are similar to those of question 2, i.e. children seemed to enjoy the avatar being there and interacting with them, regardless of the version, as we previously expected upon reading Slater et al (1999) study.

In the other hand, question 4 (Q4) and question 6 (Q6) were the ones where we found the most differences between the two groups. For instance, Q4 had an average score of 3.4 for the control version, compared with a score of 4.1 for the motivational version. It also had a lower standard deviation (SD=0.33) than the control group which suggests that most children did think they had learned something by using the system. Similarly Q6 resulted in the same pattern, but in greater differences, i.e., children in the motivational version really thought the VA helped them form the words (mean=4.2 versus mean=3.44), which goes into account with our hypothesis 5 and to the main goal of this project. For a deeper understanding of these results an independent sample t-test was performed (Table 11).

Table 11 - T-test results for questionnaire

Group Statistics

	Condition	N	Mean	Std. Deviation	Std. Error Mean
[Q2]Perceived enjoyment	Control	9	4,00	,500	,167
	Motivational	9	4,11	,782	,261
[Q4]Learning	Control	9	3,44	,527	,176
	Motivational	9	4,11	,333	,111
[Q5]Avatar liking	Control	9	4,67	,500	,167
	Motivational	9	4,78	,441	,147
[Q6]Help from avatar	Control	9	3,44	,726	,242
	Motivational	9	4,22	,441	,147

Independent Samples Test										
		Equality of Variances		t-test for Equality of Means						
						Sig. (2-tailed)	Mean Difference	Std. Error Difference	Interval of the	
		F	Sig.	t	df				Lower	Upper
[Q2]Perceived enjoyment	Equal variances assumed	3,008	,102	-,359	16	,724	-,111	,309	-,767	,545
	Equal variances not assumed			-,359	13,607	,725	-,111	,309	-,776	,554
[Q4]Learning	Equal variances assumed	11,184	,004	-3,207	16	,005	-,667	,208	-1,107	-,226
	Equal variances not assumed			-3,207	13,517	,007	-,667	,208	-1,114	-,219
[Q5]Avatar liking	Equal variances assumed	1,000	,332	-,500	16	,624	-,111	,222	-,582	,360
	Equal variances not assumed			-,500	15,754	,624	-,111	,222	-,583	,361
[Q6]Help from avatar	Equal variances assumed	2,847	,111	-2,746	16	,014	-,778	,283	-1,378	-,177
	Equal variances not assumed			-2,746	13,190	,016	-,778	,283	-1,389	-,167

We found that the most significant differences were on Q4 and Q6, which correspond to the perceived learning outcomes and the perceived help from the avatar, respectively. Question 4 for instance, had a significance level well under the 5% mark (which is the maximum threshold for significance to be verified), with 0.5% - $t(16)= -3.207$; $p=0.005$. Q6 similarly represented a significance level of around 1.5% - $t(16) = -2.746$; $p=0.014$. This suggests that the children's perception (in learning and the help they get from the agent) are much higher in the motivational version, which solidifies the idea that VA does have an effect on humans [2, 17, 22]. The following chart (Figure 30) aims to visually represent this significance found with the t-test. It tries to transmit the probability of randomly choosing a child that when asked the same two questions, their answers 'fall' between the represented mean answers (with a 95% confidence interval). In other words, if we compare the question for the perceived learning (Q4) for both versions, there is a 95% probability that the respective answers will fall between the outlined values, for each version. As we can see the difference is significant, as there is very little to no overlap between values. The same thing happens for Q6.

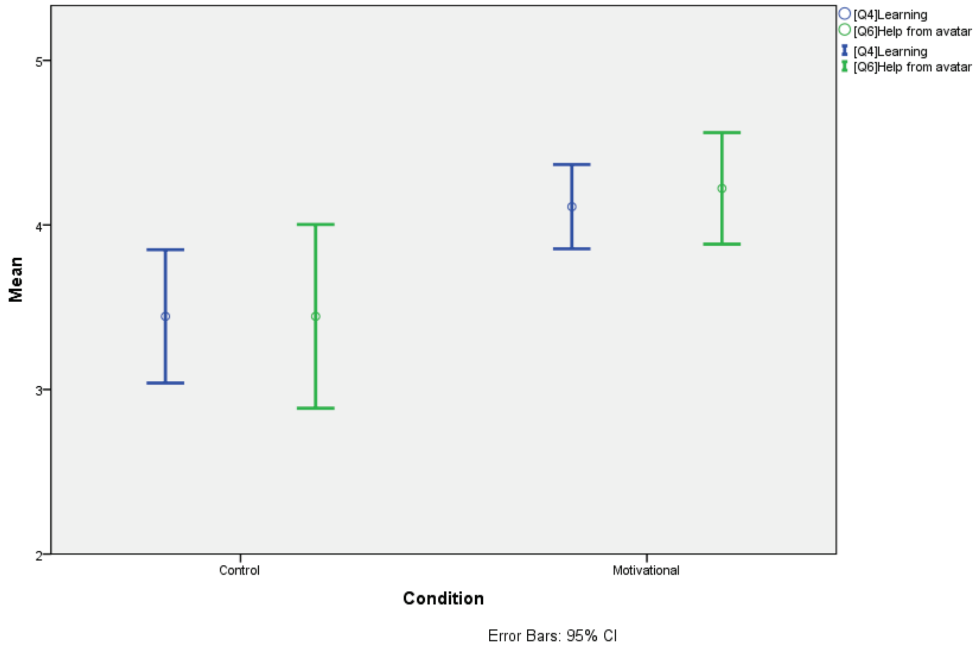


Figure 30 – Mean score for Q4 and Q6, with respective error bars

In the other hand, a more qualitative analysis told us a few things our quantitative approach couldn't.

- First we noticed that those participants that struggled a bit more than others, had a better appreciation for the VA, i.e., they paid closer attention to the virtual agent and showed excitement when they completed the word;

- We noticed that children 'dropped-out' of the game for mainly two reasons: 1) it was being too easy for them (similar with what happened in the pilot study – lack of challenge resulted in boredom state (see Csikszentmihalyi's flow model [59]); and 2) when a longer, complicated word needed to be formed. This second reason promoted mostly some frustration. These were the most frequent causers of ending the experiment, but of course there were others. For example, one child ended the participation because it had to use the bathroom, while other one simply didn't want to play anymore;

- Children also enjoyed quite a lot the visual aspect of the system, but even more, the sounds that were played when they formed the word correctly;

- Lastly, as we can verify in section 2.7.3 (related work on agent's appearance), we were expecting to see some differences in the way both genders (male and female) interacted with the system, and after an analysis of the study, we didn't confirm this. Nevertheless, upon some more research, we found that children around the ages of our participants begin to associate the male/female interaction, and prefer a same sex interaction [68], with "begin" being the keyword. Our thoughts on other reasons of why this happened, is perhaps because we had a relatively low sample or possibly the participants didn't express it out loud, but actually thought of it.

These observations showed us that a learning activity can be fun and engaging and can even be enhanced with the use of a virtual agent, as the previous work on this field of study demonstrates (Chapter 2).

As a closing remark we do believe that these results were somewhat disappointing for various reasons. One of them is that we were expecting to have a much larger difference between usage time for both systems, which is one of the ways to measure engagement, because as we said before, being engaged means being in an optimal state of flow, and when in a state of flow, time is something that passes very quickly [59]. We have no doubt that children entered the state of flow, as we observed their whole interaction and to prove that, is the result of our observation methods, where we noticed their high levels of concentration, ever since the beginning. Yet, despite all efforts to choose the images that most children of their age could recognize, sometimes they didn't, which unfortunately interrupted that same flow.

Another reason is that we believed that a longer interaction would translate into completion of more and faster tasks, however, as suggested by some researchers [41, 68], even by narrowing down to children with similar levels of knowledge, this wasn't the case. In our point of view we did not verify this, not because the knowledge levels, but mainly for the personality that they had. Personality-wise we couldn't do anything, and as a result some just wanted to complete as fast as possible, while others would take their time, independently of knowledge levels.

In our opinion, our results could've been better if we had paired the feedback with other gamification elements, such as a system that allowed children to compete with each other (noticed when we had children testing at the same time). We are well aware that it is not something we aimed at with this project, but it is something that future projects that have language learning as a goal, should implement, along with other gamification elements that we described in this thesis.

Figure 31 – Study Conduction

Chapter 5

Conclusions

The aim from project that was described so far, was to explore the importance of virtual agents in helping children in the field of language learning. After performing extensive research, it was noticed that there was a considerable amount of articles and studies, that talked about virtual agents effects on people (children to adults), that could vary from motivating to perform physical exercise [17], to playing and learning within a virtual world [18]. However systems that target language learning specifically with virtual agents to provide feedback (play the role of a teacher), from what we could find, is still in its infancy.

With this in mind, we had as a goal to create a fun system that was able to create a motivational environment (as described in Chapter 1), so children could be engaged for longer periods of time. By studying how to engage children and make activities more fun/engaging (Chapter 2) we found that what most children like when learning is the ability of being 'independent'. They like to guide their activities by their own, and only when necessary is that they need an adult. This is why we designed the virtual agent to encourage them on continuing their task even when they weren't being successful. This resulted on a system that ran in an android tablet, and was only possible by studying what children liked. To do this, what better way than involving children in the design process? By asking a couple of children if they 'liked this' or 'how would you prefer this' types of questions, we managed to create a system that was visually appealing and that ultimately led us to our pilot study. After the visuals of the system were created, only a few minor alterations were necessary to complete it (graphics related) and the proof of this is the final design that was nothing like the initial design for the system (see Figure 32 and Figure 33). Afterwards, we proceeded to think about and create the most important aspect of the project - the virtual agent.

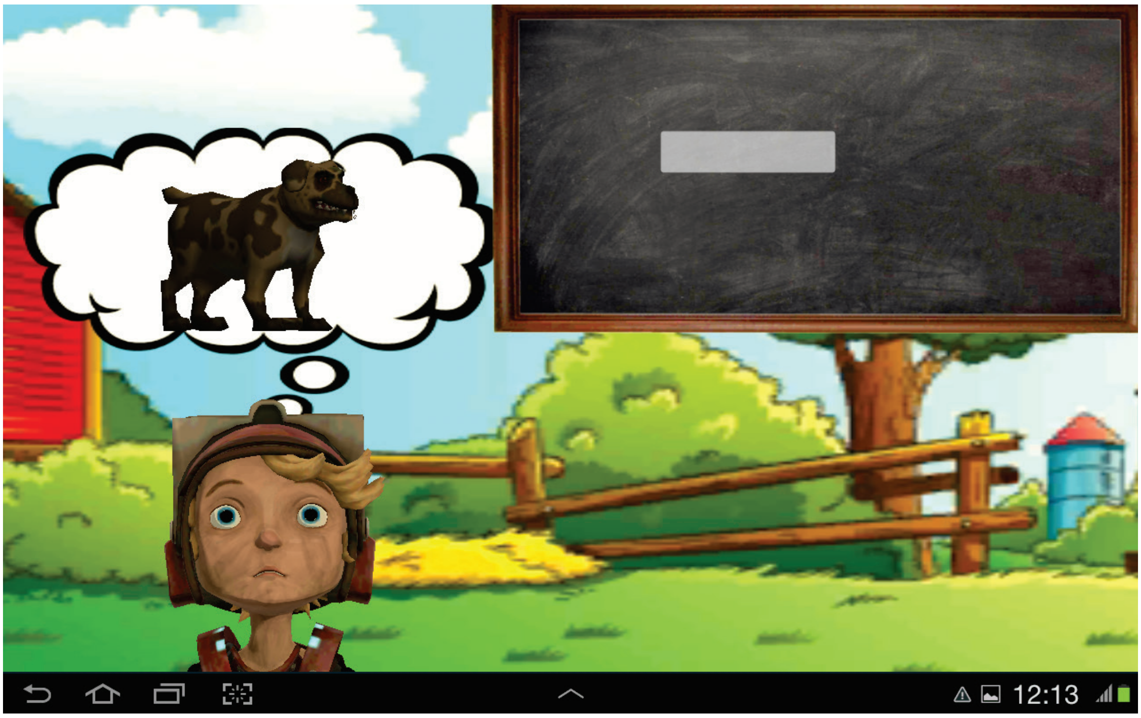


Figure 32 – Initial main screen

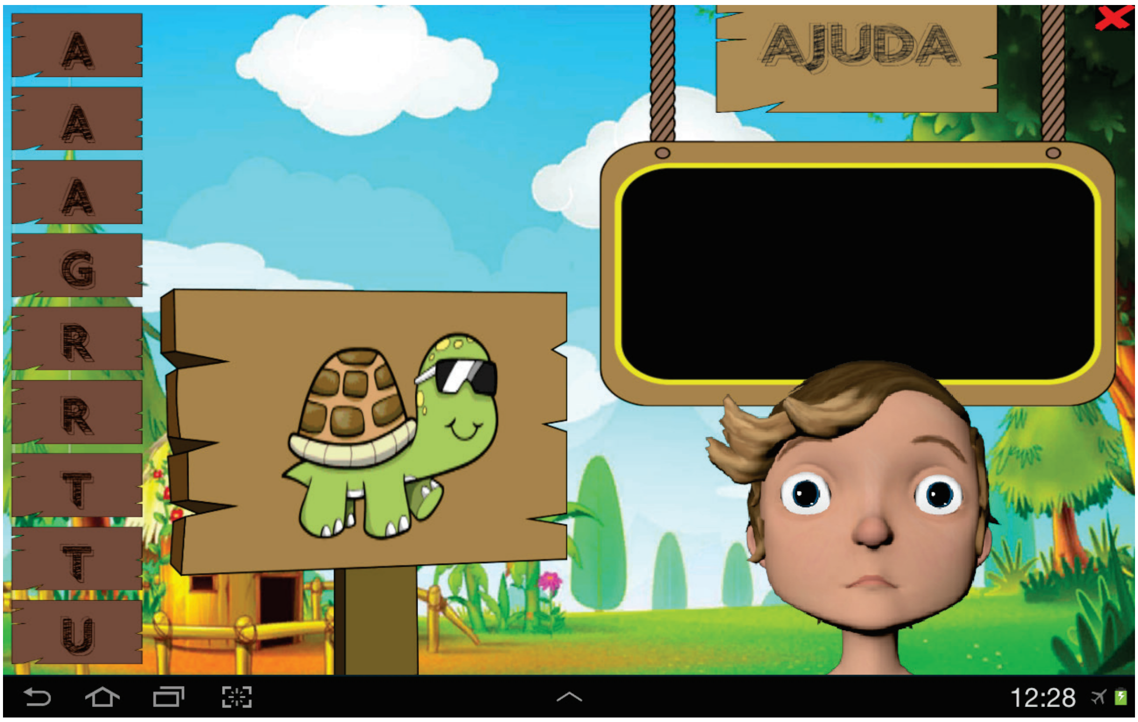


Figure 33 – Final main screen

By using this approach we know we hit the right point for a good representation of the system, along with the VA to enhance the learning activity. Also, we found children to be alert to the VA, and so, we can see these types of system eventually becoming successful among children for language learning.

Concerning the prototype, it is important to notice, that we had a much more beneficial response from children for the motivational feedback provided in this final version, when compared to the pilot study. As we explained previously we found that flow was being disturbed by the way feedback was being given. The solution we found consisted on playing the same animation on a separate screen. However we had to create another application - a 'desktop' application, because android didn't support the feature we wanted (playing in a different/separate screen). This application was constantly listening to certain 'flags' that were broadcasted by the tablet, such as "strike", and would then play in the computer screen an animation related to the "strike" event. By providing motivational feedback in this fashion, we successfully removed the factor that sometimes provided some frustration to children. This in our opinion contributed to a more pleasant experience for the study.

With this said, we should address the main research question that guided us throughout our work. In the next section we answer and summarize about the question "How can we engage children in language learning by delivering feedback through virtual characters?", and in this way, complement the work that is described throughout the document, from the literature review to the results found with the elaboration of the study.

5.1. Research question explained

5.1.1. How can we engage children in language learning by delivering feedback through virtual characters?

Ever since the beginning we knew that virtual agents could be a tricky thing to do, because of the complexity that is to 'create' something that behaves in a similar way to humans. Upon reviewing the importance of virtual characters (see section 2.6), we noticed that delivering feedback in the correct way is more difficult than one might think. There are too many aspects inherent to them (VA's) that contribute each in its own way, for engagement. The first factor is, for instance the fact that the virtual character appearance, can make it be more believable, or even the gender, since usually, depending on age range, children behave differently when interacting with other children of different genders. The second factor is that the way that feedback is delivered, is tightly paired with the manner that the character provides it – either the voice tone or the facial expressions. For instance in Chapter 2 we provided the example that when a negative message is delivered with a smile, the child might understand it as a positive thing, when it is not the case, which causes them to learn incorrectly.

In sum and in our point of view after performing this study, the way in which the virtual character can engage children, is by following a few general guidelines about them, where most of them are related to the appearance, behaviour and the timing for when intervene (provide feedback):

- For starters, we should use a virtual agent that is age appropriate to the target population of the system;
- Choose a virtual agent that has an appealing appearance, i.e. a friendly appearance;
- Make them behave according to its appearance (in a way give some personality that relates to their looks, for example, an energetic person versus a more calmer, quiet one);
- Deliver rich feedback, i.e. feedback that is related with something the child did, and not some meaningful 'pre-defined' feedback;
- Based on our experience, make the character behave without disturbing the flow (without disturbing the execution of the task);
- Intervene in the correct time it is needed. This contributes to the above recommendation of not disturbing flow.

- About the system in general, try to incorporate some children to help in the evolution of the design, so that we can have an opinion from someone who 'understands' what they're talking about;

- Lastly, use plenty of music and sound effects, because children really enjoy them, as it was shown in the conduction of the field study.

As we saw in our results, by following these steps to create a VA, we can obtain a working system that can help towards the future of education, not only in language learning, but in all other disciplines.

5.2. Limitations

Our project suffered some limitations concerning our testing phase, which was consequently reflected in our results. First, when the prototype was finished we found ourselves in a holiday season, which compromised our testing capabilities, because our idea was to test the prototype with children from two different classrooms of a local school. However due to bureaucracy reasons we weren't allowed to test our system with a local classroom, which put us at risk of not having any user tests, and thus an incomplete work.

With this in mind, we needed urgently to find an alternative, and the alternative found (as mentioned in Chapter 4) was first by finding a couple of children, and then talking to their friends (and parents) and so on. Only this way, we managed to test the system, but with the drawback that our population sample wouldn't be nearly vast enough as if it was in the school, and therefore limiting the results we obtained.

5.3. Obstacles faced

This project was initially thought to have a different path than what it actually had. First it was thought to be made with a tangible user interface (TUI), and a platform was actually designed and constructed. The platform was where the tablet would be positioned and the word formation would be made possible by placing small cubes (5x5 cm) (where each corresponded to a letter of the alphabet) in it. The cubes had an LED inside that light up red or green, according to the placement of the letter. What happened was that, that same platform was literally destroyed by someone at the 'workplace' where it was stored, and made it impossible to redo it with the time that was left for the deadline. The solution to overcome this major issue, was the one that was described ever since the beginning of this document (making the 'physical' become 'virtual').

For the development of the system, the first obstacle that was encountered was with the integration of the virtual agent software editor with the program where the system was developed (Unity). In other words, the plugin supplied with the virtual character for unity had issues when trying to create a connection to the internet which didn't allow the use of the, initially defined, virtual agent software.

After finding another suited software for the creation of the virtual agent (CrazyTalk7), we found yet another problem that is somewhat similar to the one before. When downloading the program from its webpage we noticed that it offered a pro version trial with the duration of fifteen days. Since it was a limited license the thought here was to finish all animations for the expressions and only then activate the pro version and make the import to unity, so we could work on them from there. After finishing the editing of these, what happened is that the pro version of the software only allowed with a predefined template. By then we had hit a brick wall, because that was the way it was supposed to be made. After some thinking and investigation about Unity, we found a feature that allowed to use video as a texture, instead of fixed images, so it was decided that was how it was going to be done. The small problem that was noticed is that the video had a white background and so we spent some time in video editing software to remove it, but when importing to unity, the video file format that unity supported didn't support transparency. What we did was work on a background that would merge with the system background and hope it didn't become very noticeable. With this done, and after doing a couple functional animations, we wanted to experiment in the tablet how it ran, and to our surprise (again) it didn't work. After some digging around, we found that android did not support the use of video textures, which threw us back to the beginning. This time we decided to lookup something that was compatible and found the "Sprites" feature built into unity. This option worked great but there was a problem. A single animation with just a few seconds length, could use over 100Mb in size, which made the app run slowly and occupy a lot of storage, which wasn't practical. The solution was to create a combination of this and a video playback, which can be seen in the final system. It didn't necessarily remained as good as we wanted it to be, but it was the only way found.

5.4. Future Work

Upon finishing our work, we can say that there is still room for improvement. For instance future work could work on exploring the gazes conveyed by the virtual character, that wasn't possible in this project due to software restrictions. It's a very important feature, because research has shown that gaze is related to task performance, and when children are gazed at, they tend to recall the information better than those who are not [46]. Gaze is likewise important when conveying facial expressions – for instance when the agent is thinking, 'looking up' is a universal way to show that some cognitive process is taking place.

Also it would be interesting if children could have the ability of choosing between some different virtual agents (male or female), or even 'create' one, where they could identify with or relate with. This feature despite contributing to maintain engagement with children, also contributes to the concept of gamification, where one of the elements is precisely to let children choose an avatar.

Another option would be to design these types of system in a way that allows teachers to add their own objects and respective sound effects. This wasn't done in this project since it was about a more exploratory project that aimed to see how these types of systems behave in real situations, and thus, future work based could and should have this feature built-in.

As stated previously this project was about getting insights on the effect that the VA's feedback has on children, which alone is not a very complex gamified activity. However for future work that wants to have a complete experience of gamification (i.e. adding more gamification elements), we suggest a creation of a game with levels where difficulty increases steadily according to player skills (like stated in flow theory, to prevent loss of interest) and also consider the possibility of adding some small in-game rewards, such as a piece of clothing for their avatar, etc.)

On a final note, we do believe to have contributed positively for the area of education by using the technology we currently dispose of, technology itself that has been being proven numerous time in the last years, as being a ‘plus’ in education whether for language learning or for other ends. This thesis allowed to add a new project in the few that are available on this area, for future research, and proved that gamification (feedback specifically) is becoming more and more important as we can verify by the number of studies they have influence on, and are slowly becoming as an alternative to teach young children.

References

1. Ryokai, K., Vaucelle, C., & Cassell, J. (2003). Virtual peers as partners in storytelling and literacy learning. Journal of computer assisted learning, 19(2), 195-208.

2. Zibrek, K., Hoyet, L., Ruhland, K., & McDonnell, R. (2013). Evaluating the effect of emotion on gender recognition in virtual humans. Proceedings of the ACM Symposium on Applied Perception. ACM.

3. McDonnell, R., Jörg, S., McHugh, J., Newell, F., & O'Sullivan, C. (2008). Evaluating the emotional content of human motions on real and virtual characters. Proceedings of the 5th symposium on Applied perception in graphics and visualization. ACM.

4. Kasap, Z., Ben Moussa, M., Chaudhuri, P., & Magnenat-Thalmann, N. (2009). Making them remember—emotional virtual characters with memory. Computer Graphics and Applications, IEEE, 29(2), 20-29.

5. Liu, Z., & Lu, Y. (2008). A motivation model for virtual characters. Machine Learning and Cybernetics, 2008 International Conference on. IEEE.

6. Vala, M., Paiva, A., & Prada, R. (2004). From motion control to emotion influence: Controlling autonomous synthetic characters in a computer game. Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 3. IEEE Computer Society.

7. Zalewski, L., & Gong, S. (2005). 2d statistical models of facial expressions for realistic 3d avatar animation. Computer Vision

and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE.

8. Cootes, T. F., Edwards, G. J., & Taylor, C. J. (1998). Active appearance models. Computer Vision—ECCV'98, 484-498.

9. Conte, R., & Castelfranchi, C. (1995). Cognitive and social action. Psychology Press.

10. Ortony, A., Clore, G. L., & Collins, A. (1990, May 25). The cognitive structure of emotions. Cambridge university press.

11. McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. Journal of personality, 60(2), 175-215.

12. Argyle, M. (2013, April 15). Bodily communication. Routledge.

13. McDonnell, R., & O'Sullivan, C. (2010). Movements and voices affect perceived sex of virtual conversers. Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization. ACM.

14. Tinwell, A., Grimshaw, M., Nabi, D. A., & Williams, A. (2011). Facial expression of emotion and perception of the Uncanny Valley in virtual characters. Computers in Human Behavior, 27(2), 741-749.

15. Gulz, A. (2004). Benefits of virtual characters in computer based learning environments: Claims and evidence. International Journal of Artificial Intelligence in Education, 14(3), 313-334.0

16. Liu, Z., & Pan, Z. G. (2005). An emotion model of 3d virtual characters in intelligent virtual environment. Affective Computing and Intelligent Interaction, 629-636.

17. Baylor, A. L. (2009). Promoting motivation with virtual agents and avatars: role of visual presence and appearance. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1535), 3559-3565.

18. Merchant, G. (2009). Literacy in virtual worlds. Journal of Research in Reading, 32(1), 38-56.

19. McQuiggan, S. W., Rowe, J. P., & Lester, J. C. (2008). The effects of empathetic virtual characters on presence in narrative-centered learning environments. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM.

20. Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. Presence: Teleoperators and virtual environments, 7(3), 225-240.

21. André, E., Klesen, M., Gebhard, P., Allen, S., & Rist, T. (2000). Integrating models of personality and emotions into lifelike characters. Affective interactions, 150-165.

22. Vinayagamoorthy, V., Gillies, M., Steed, A., Tanguy, E., Pan, X., Loscos, C., et al. (2006). Building expression into virtual characters.

23. Prendinger, H., Mori, J., & Ishizuka, M. (2005). Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game. International journal of human-computer studies, 62(2), 231-245.

24. Slater, M., Pertaub, D., & Steed, A. (1999). Public speaking in virtual reality: Facing an audience of avatars. Computer Graphics and Applications, IEEE, 19(2), 6-9.

25. Chittaro, L., & Serra, M. (2004). Behavioral programming of autonomous characters based on probabilistic automata and personality. Computer animation and virtual worlds, 15(3-4), 319-326.

26. Ochs, M., Sabouret, N., & Corruble, V. (2009). Simulation of the dynamics of nonplayer characters' emotions and social relations in games. Computational Intelligence and AI in Games, IEEE Transactions on, 1(4), 281-297.

27. Lee, J. J., & Hammer, J. (2011). Gamification in education: What, how, why bother?. Academic Exchange Quarterly, 15(2), 146.

28. Simões, J., Redondo, R. D., & Vilas, A. F. (2013). A social gamification framework for a K-6 learning platform. Computers in Human Behavior, 29(2), 345-353.

29. Cheong, C., Cheong, F., & Filippou, J. (2013). Quick Quiz: A Gamified Approach for Enhancing Learning. PACIS.

30. Brewer, R., Anthony, L., Brown, Q., Irwin, G., Nias, J., & Tate, B. (2013). Using gamification to motivate children to complete empirical studies in lab environments. Proceedings of the 12th International Conference on Interaction Design and Children. ACM.

31. Domínguez, A., Saenz-de-Navarrete, J., De-Marcos, L., Fernández-Sanz, L., Pagés, C., & Martínez-Herráiz, J. (2013). Gamifying learning experiences: Practical implications and outcomes. Computers & Education, 63, 380-392.

32. Stott, A., & Neustaedter, C. (2013). Analysis of gamification in education. Surrey, BC, Canada.

33. Kapp, K. M. (2012). Games, gamification, and the quest for learner engagement. T+ D, 66(6), 64-68.

34. Ronimus, M., Kujala, J., Tolvanen, A., & Lyytinen, H. (2014). Children's engagement during digital game-based learning of reading: The effects of time, rewards, and challenge. Computers & Education, 71, 237-246.

35. Lovio, R., Halttunen, A., Lyytinen, H., Näätänen, R., & Kujala, T. (2012). Reading skill and neural processing accuracy improvement after a 3-hour intervention in preschoolers with difficulties in reading-related skills. Brain research, 1448, 42-55.

36. Andrist, S., Pejsa, T., Mutlu, B., & Gleicher, M. (2012). Designing effective gaze mechanisms for virtual agents.

Proceedings of the SIGCHI conference on Human factors in computing systems. ACM.

37. Kapp, K. M. (2012, May 1). The gamification of learning and instruction: game-based methods and strategies for training and education. John Wiley & Sons.

38. Roussou, M. (2004). Learning by doing and learning through play: an exploration of interactivity in virtual environments for children. Computers in Entertainment (CIE), 2(1), 10-10.

39. Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining gamification. Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. ACM.

40. LIVRO play and learning in the early days

41. Lucero, A., Karapanos, E., Arrasvuori, J., & Korhonen, H. (2014). Playful or Gameful?: creating delightful user experiences. interactions, 21(3), 34-39.

42. Barblett, L. (2010). Why play-based learning?. 16(3), 4.

43. Department for Children, Schools and Families, “Learning, Playing and Interacting.” Sep-2009.

44. Otteson, J. P., & Otteson, C. R. (1980). Effect of teacher's gaze on children's story recall. Perceptual and Motor Skills, 50(1), 35-42.

45. Pan, Q. (2014). Nonverbal Teacher-student Communication in the Foreign Language Classroom. Theory and Practice in Language Studies, 4(12), 2627-2632.

46. Mutlu, B., Forlizzi, J., & Hodgins, J. (2006). A storytelling robot: Modeling and evaluation of human-like gaze behavior. Humanoid Robots, 2006 6th IEEE-RAS International Conference on. IEEE.

47. Tunstall, P., & Gsipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. British Educational Research Journal, 22(4), 389-404.

48. National Association for the Education of Young Children, “Learning to Read and Write: Developmentally Appropriate Practices for Young Children.” May-1998. Learning to Read and Write

49. Tunstall, P., & Gsipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. British Educational Research Journal, 22(4), 389-404.

50. Cotton, K. (1988, May). Instructional reinforcement. Northwest Regional Educational Laboratory.

51. Askew, S. (2000). Feedback for learning (S. Askew). Psychology Press.

52. R. Coe, C. Aloisi, S. Higgins, and L. E. Major, “What makes great teaching? Review of the underpinning research.,” 2014. What makes great teaching?

53. Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning, and the psychological processes that mediate their relationship. Journal of educational psychology, 94(3), 545.

54. Prensky, M. (2003). Digital game-based learning. Computers in Entertainment (CIE), 1(1), 21-21

55. Benabou, R., & Tirole, J. (2003). Intrinsic and extrinsic motivation. The Review of Economic Studies, 70(3), 489-520.

56. Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. Contemporary educational psychology, 25(1), 54-67.

57. Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. Journal of personality and Social Psychology, 18(1), 105.

58. Csikszentmihalyi, M., & Csikzentmihaly, M. (1991, March). Flow: The psychology of optimal experience (Vol. 41). New York: HarperPerennial.

59. Nakamura, J., & Csikszentmihalyi, M. (2002). The concept of flow. Handbook of positive psychology, 89-105.

60. Taylor, C. (2012). Improving attendance at school.

61. Reinking, D., McKenna, M. C., Labbo, L. D., & Kieffer, R. D. (1998, April 1). Handbook of literacy and technology: Transformations in a post-typographic world (D. Reinking, M. C. McKenna, L. D. Labbo, & R. D. Kieffer). Routledge.

62. Strickland, D. S., & Morrow, L. M. (1989). Emerging literacy: Young children learn to read and write.. International Reading Association, 800 Barksdale Rd., PO Box 8139, Newark, DE 19714-8139 (No. 352, $10.00 member, $15.00 nonmember).

63. Alexiu, T. M., & Sordé, T. (2011). How to turn difficulties into opportunities: Drawing from diversity to promote social cohesion. International Studies in Sociology of Education, 21(1), 49-62.

64. Frick, J. R., & Wagner, G. G. (2001). Economic and social perspectives of immigrant children in Germany (No. 301). IZA Discussion paper series.

65. L. D. Labbo, M. S. Love, and T. Ryan, “A Vocabulary Flood: Making Words ‘Sticky’ With Computer-Response Activities,” The Reading Teacher, vol. 60, no. 6, pp. 582–588, Mar. 2007.

66. LEMKE, JL. "Metamedia Literacy: Transforming Meanings And Media." 2011.

67. Labbo, Linda D, and David Reinking. "Negotiating the multiple realities of technology in literacy research and instruction." Reading Research Quarterly 34.4 (1999): 478-492.

68. P. Markopoulos, J. C. Read, S. MacFarlane, and J. Hoysniemi, Evaluating Children’s Interactive Products: Principles and Practices for Interaction Designers. Morgan Kaufmann, 2008.

69. Liu, Min. "An Exploratory Study of How Pre-Kindergarten Children Use the Interactive Multimedia Technology: Implications for Multimedia Software Design." (1996).

70. K. Magnuson;, C. Lahaie;, and J. Waldfogel, “Preschool and School Readiness of Children of Immigrants.” ProQuest Education Journals, Dec-2006.

71. Christie, Frances, and Ray Misson. Literacy and schooling. Psychology Press, 1998.

72. Magnuson, Katherine, Claudia Lahaie, and Jane Waldfogel. "Preschool and School Readiness of Children of Immigrants*." Social Science Quarterly 87.5 (2006): 1241-1262.

73. "Early Literacy Skills, Behaviors and Importance - Maine.gov." 2012

74. Romero, M. "A National Portrait of Chronic Absenteeism in the Early Grades." 2007.

75. Daraganova, Galina. "Is it OK to be away?” (2012).

76. Attendance in the Early Grades: - Attendance Works." 2014.

77. Skinner, Burrhus Frederic. "The behavior of organisms: An experimental analysis." (1938).

78. H. L. O’Brien and E. G. Toms, “What is user engagement? A conceptual framework for defining user engagement with technology,” Journal of the American Society for Information Science and Technology, vol. 59, no. 6, pp. 938–955, Apr. 2008.

79. Newmann, F. M. (1992). Student engagement and achievement in American secondary schools. Teachers College Press, 1234 Amsterdam Avenue, New York, NY 10027

80. Strasburger, V. C., Hogan, M. J., Mulligan, D. A., Ameenuddin, N., Christakis, D. A., Cross, C., et al. (2013). Children, adolescents, and the media. Pediatrics, 132(5), 958-961.

81. Marco, J., Cerezo, E., Baldassarri, S., Mazzone, E., & Read, J. C. (2009). Bringing tabletop technologies to kindergarten children. Proceedings of the 23rd British HCI group annual conference on people and computers: Celebrating people and technology. British Computer Society.

82. Rosenberg-Kima, R., Baylor, A. L., Plant, E. A. & Doerr, C. 2007 The importance of interface agent visual presence: voice alone is less effective in impacting young women’s attitudes toward engineering. Persuasive 2007, vol. 4744., pp 214–222.

83. Baylor, A. L. & Plant, E. A. 2005 Pedagogical agents as social models for engineering: the influence of appearance on female choice. In Artificial intelligence in education: supporting learning through intelligent and socially informed technology, vol. 125 (eds C.-K. Looi, G. McCalla, B. Bredeweg & J. Breuker), pp. 65–72. Bristol, UK: IOS Press.

84. Bailenson, J. N., Beall, A. C., Blascovich, J., Raimundo, M., & Weisbuch, M. (2001). Intelligent agents who wear your face: Users' reactions to the virtual self. Intelligent Virtual Agents. Springer Berlin Heidelberg.

85. Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. International Journal of Artificial intelligence in education, 11(1), 47-78.

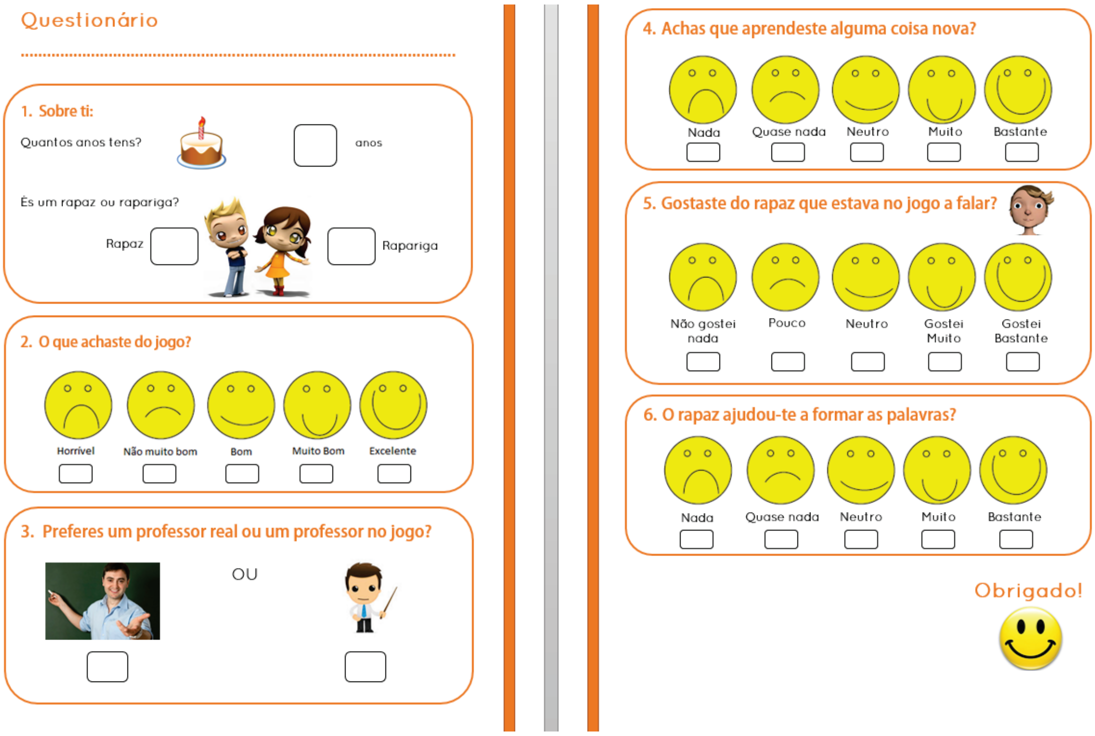
Appendix A

Questionnaire answers

The table below (Table 12) represents the raw data of the questionnaire answers for each group of participants. Q2, Q4, Q5, Q6 and Q3 refer to the number of the question in the questionnaire handed to children in the end of their experiment. The numbers in red represent the excluded questionnaires due to low literacy levels.

Table 12 – Summarized answers of questionnaire

	Participant	Q2	Q4	Q5	Q6	Q3
Control group	**1**	4	3	5	3	Virtual
	2	4	3	5	3	Real
	3	4	4	5	5	Virtual
	4	4	3	4	3	Real
	5	4	3	5	4	Virtual
	6	4	3	5	3	Virtual
	7	3	4	5	4	Virtual
	8	5	4	4	3	Real
	9	4	4	4	3	Virtual
Motivational group	**10**	4	5	5	4	Virtual
	11	4	4	5	5	Virtual
	12	3	4	5	4	Real
	13	5	4	5	4	Virtual
	14	4	4	4	5	Virtual
	15	3	4	5	4	Real
	16	5	4	5	4	Virtual
	17	5	4	4	4	Real
	18	4	4	5	4	Virtual
		1	2	3	1	Real
		2	1	2	1	Real
					Total Real:	6
					Total Virtual:	12


Questionário

1. Sobre ti:
Quantos anos tens? anos
És um rapaz ou rapariga?
Rapaz Rapariga

2. O que achaste do jogo?
Horrível Não muito bom Bom Muito Bom Excelente

3. Preferes um professor real ou um professor no jogo?
OU

4. Achas que aprendeste alguma coisa nova?
Nada Quase nada Neutro Muito Bastante

5. Gostaste do rapaz que estava no jogo a falar?
Não gostei nada Pouco Neutro Gostei Muito Gostei Bastante

6. O rapaz ajudou-te a formar as palavras?
Nada Quase nada Neutro Muito Bastante

Obrigado!

Figure 34 - Questionnaire