



TÉCNICO
LISBOA



Virtual Tutor

Virtual Tutoring Agent using Empathy and Rapport Techniques

André Guilherme Carrasqueira Lima

Thesis to obtain the Master of Science Degree in

Information Systems and Computer Engineering

Supervisor: Prof./Dr.Carlos António Roque Martinho

Examination Committee

Chairperson: Prof. António Manuel Ferreira Rito da Silva

Supervisor: Prof./Dr.Carlos António Roque Martinho

Member of the Committee: Prof. Ana Paula Boler Cláudio

September 2017

The computer is incredibly fast, accurate, and stupid. Man is unbelievably slow, inaccurate, and brilliant. The marriage of the two is a challenge and opportunity beyond imagination.

Stuart G. Walesh

Acknowledgments

I would like to thank Prof./Dr.Carlos António Roque Martinho and Ricardo Eugénio Proença Rodrigues for all the patience, meetings, explanations, e-mails and above all, the availability for guiding, helping and clearing any and all questions that would originate in my mind.

I would also like to thank all the committee of the "*Tutoria Virtual*" program made up of members from "*Universidade Aberta*" and "*Faculdade de Ciências da Universidade de Lisboa*", for the time and taken to analyze and the input on our project that made it be an even better, more functioning prototype.

Last but not least, I would like to thank my family and friends, and everyone who supported me through this process and helped me gain courage on this hard but rewarding road for the conclusion of my masters.

Abstract

Going to University is a big change for someone and the student normally enters University feeling motivated. However, when the difficulties start to arrive and the reality that the effort required to succeed is much bigger than on the previous steps of their academic lives starts dawning on the students that just entered University they tend to lose motivation and cannot find the right methods of dealing with these loss of motivation, some of them fail courses they would on other circumstances find easy. Some universities have decided to implement a tutoring program in which the students are accompanied by a real tutor of the faculty. We intend to add to this program by delivering a tutor that is always available for the student and will help them emotionally. We brought the metaphor of two tutor characters with complimentary personalities to accompany the student throughout their semester. We studied what factors will help the student's emotional state and we arrived at rapport, empathy and relationship. We did an exploratory work on which one was more important by showing different behaviors for a virtual tutor agent in an app environment. This work was partially supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with ref. UID/CEC/50021/2013, and FCT grant from project Tutoria Virtual with ref. TDC/IVC-PEC/3963/2014

Keywords

Empathy, Rapport, Relationships, Virtual Tutors, Education, Emotional Intelligence, Technology, Virtual Agents, Virtual Assistants, Future Education.

Contents

1	Introduction	xv
1.1	Motivation	xvi
1.1.1	Problem	xvi
1.1.2	Hypothesis	xvi
1.2	Original Contributions	xvii
1.3	Thesis Outline	xvii
2	State of the art and Related Work	xix
2.1	Virtual Tutoring Agents	xx
2.2	Relationships	xxi
2.3	Rapport	xxiv
2.4	Empathy	xxvii
3	The Virtual Tutor app	xxxix
3.1	Approaching the Concept	xxxix
3.1.1	The Quadrant System	xxxix
3.1.2	Impacts of confidence, motivation and importance	xxxix
3.1.3	Normal scenario of interaction	xxxix
3.1.4	Managing Emotions	xxxix
3.1.5	Design and Interactivity	xxxix
3.1.6	Dialogue	xxxix
3.2	Usability User Tests	xxxix
3.2.1	Presenting the first functional prototype	xxxix
3.2.2	The MoJo	xxxix
3.2.3	Setting up the experiment	xxxix
3.2.4	Results	xxxix
3.3	The final Prototype	xl
3.3.1	Dialogue	xl
3.3.2	Checkpoints	xl
3.3.3	Sprites and Expressions	xl
3.3.4	Integration	xl
3.3.5	Design	xl

3.3.6	Sound	xliv
3.3.7	Architecture	xliv
3.3.8	Class Diagram	xlvi
3.3.9	Concluding remarks	xlvi
4	The Experiment	xlix
4.1	Experimental Protocol	i
4.1.1	Relationship stages - Acquaintance and Maintenance	i
4.1.2	Rapport	li
4.1.3	Empathy - Posture mimicry and Perspective taking	lii
4.1.4	Procedures	liii
5	Results from the Experiment	iv
5.1	Participants	lvi
5.1.1	Age	lvi
5.1.1.A	Gender	lvi
5.1.1.B	Student	lvi
5.1.1.C	Participants' other demographic data	lvi
5.2	Results of the experiment	lvii
5.2.1	Importance	lvii
5.2.2	Adjectives	lviii
5.3	Statistic Tests and respective results	lix
5.3.1	Shapiro-Wilk test to check normality	lix
5.3.2	Friedman General test	lx
5.3.3	Wilcoxon signed-rank test in pairs	lx
5.3.4	Friedman test with a split the dataset	lx
5.3.5	Discussion of Results	lxii
6	Conclusions and Future Work	lxv
	Bibliography	lxvii
	Appendix A The Questionnaire	A-1
	Appendix B Demographic statistic results	B-1
	Appendix C Lickert Scales	C-1
	Appendix D Adjectivation statistic results	D-1
	Appendix E Shapiro-Wilk and Freidman test results	E-1
	Appendix F Wilcoxon Signed rank test	F-1

List of Figures

2.1	[1]'s ERISA's framework overview	xxiv
2.2	[2]'s Goal Tree of the Rapport Model	xxvi
2.3	I. Leite's iCat interacting	xxviii
2.4	Maxx and Emma's interaction	xxix
3.1	Example of the four quadrants for student's motivation and confidence	xxxiii
3.2	The two tutor agents in our prototype	xxxv
3.3	Example of the dialogue in Oxenfree	xxxvi
3.4	Example of the timeout scenario in the dialogue	xxxvii
3.5	Mojo Participation	xxxix
3.6	The experiment's results	xxxix
3.7	Example of Checkpoints in the UI	xli
3.8	The final Avatars	xlii
3.9	Each agent's architecture	xliv
3.10	Virtual Tutor's class diagram	xlvi
4.1	Example of one of the clips interactions with a cursor.	li
5.1	Statistic information on the Importance variable in the six videos	lviii
B.1	Participants' Age Statistics	B-2
B.2	Participants' gender statistics	B-2
B.3	Pie chart with the participants' gender	B-2
B.4	Participants' student information statistics	B-3
B.5	Pie chart with the participants' student information	B-3
B.6	Pie chart with the participants' information on the usage of mobile apps on their smart- phone	B-3
B.7	Pie chart with the participants' information on the usage of the calendar apps on their smartphone	B-4
B.8	Pie chart with the participants' information on the familiarity with Virtual Assistants	B-4
B.9	Pie chart with the participants' information on the preference to interact with NPC's in games	B-5

C.1	Frequency table regarding the importance of the video focusing on the acquaintance level of the relationship	C-2
C.2	Frequency table regarding the importance of the video focusing on the maintenance level of the relationship	C-2
C.3	Frequency table regarding the importance of the video focusing on the rapport strategy of self disclosure for positivity	C-2
C.4	Frequency table regarding the importance of the video focusing on the rapport strategy of back channeling for coordination and mutual attention	C-3
C.5	Frequency table regarding the importance of the video focusing on the empathy strategy of posture mimicry	C-3
C.6	Frequency table regarding the importance of the video focusing on the empathy strategy of perspective taking	C-3
D.1	Statistic information on the number positive adjectives used to describe each tutor . . .	D-2
D.2	Statistic information on the number negative adjectives used to describe each tutor . .	D-2
D.3	Statistic information on the number times each tutor was qualified as showing emotions	D-2
D.4	Statistic information on the number times each tutor was qualified as being serious . . .	D-2
D.5	Statistic information on the number times each tutor was qualified as being friendly . . .	D-2
D.6	Statistic information on the number times each tutor was qualified as being supportive or helpful	D-2
E.1	Shapiro-Wilk's results on the importance of the acquaintance phase of the relationship in the Tutors	E-2
E.2	Shapiro-Wilk's results on the importance of the maintenance phase of the relationship in the Tutors	E-2
E.3	Shapiro-Wilk's results on the importance of the rapport strategy of self disclosure for the increase of positivity in the Tutors	E-2
E.4	Shapiro-Wilk's results on the importance of the rapport strategy of back channeling for the increase of coordination and mutual attention in the Tutors	E-2
E.5	Shapiro-Wilk's results on the importance of the empathy generation strategy of posture mimicry in the Tutors	E-2
E.6	Shapiro-Wilk's results on the importance of the empathy generation strategy of perspective taking in the Tutors	E-2
E.7	Friedman test results on the variance of importances of all behaviors shown in the videos	E-3
F.1	Matrix of significance of all the pairs of Wilcoxon signed rank tests performed	F-2
G.1	Results of the Friedman test counting only participants of gender male	G-2
G.2	Results of the Friedman test counting only participants of gender male	G-3
G.3	Results of the Friedman test counting only participants of age 24	G-4
G.4	Results of the Friedman test counting only participants of age higher than 24	G-5

G.5	Results of the Friedman test counting only participants of age lower than 24	G-6
G.6	Results of the Friedman test counting only participants that always use their calendar app	G-7
G.7	Results of the Friedman test counting only participants that never use their calendar app	G-8
G.8	Results of the Friedman test counting only participants that occasionally use their calendar app	G-9
G.9	Results of the Friedman test counting only participants that prefer games where they have to interact with a NPC	G-10
G.10	Results of the Friedman test counting only participants that prefer games where they are not forced to interact with NPC	G-11
G.11	Results of the Friedman test counting only participants that have no preferences about the presence of NPCs in a game	G-12
G.12	Results of the Friedman test counting only participants that use mobile apps everyday .	G-13
G.13	Results of the Friedman test counting only participants that do not use mobile apps everyday	G-14
G.14	Results of the Friedman test counting only participants that are university students . . .	G-15
G.15	Results of the Friedman test counting only participants that are not university students .	G-16
G.16	Results of the Friedman test counting only participants that never had interacted with a Virtual Assistant	G-17
G.17	Results of the Friedman test counting only participants that had occasionally interacted with a Virtual Assistant	G-18

1

Introduction

Contents

1.1 Motivation	xvi
1.2 Original Contributions	xvii
1.3 Thesis Outline	xvii

1.1 Motivation

It is part of a student's life to find themselves with problems of adaptation to their new Academic life when they reach University. To counteract this effect, some universities started having more office hours with the course professors so that students feel motivated to ask questions and clear all misunderstandings and questions that came up during that week. Some universities also started a tutoring program in which one professor is in charge of a group of students. This professor is normally not from any course these students are attending but he is tasked with keeping their morale up and checking up on them in case any of the above mentioned problems arise.

As we come into a more digital age where everybody is connected to each other and using a smartphone is part of their daily routine by creating an app that would coach the student, a sorts of "Virtual Tutor" that would always be there for the student for the more daily tasks , we hope to improve the tutoring program.

Our Virtual Tutor is inserted in the "*Tutoria Virtual*" research. A program from "*Fundação para a Ciência e Tecnologia*", a Portuguese public agency for science and technology. This program studies the pedagogical impact of Encapsulated Virtual Agents in online learning environments as Virtual Tutors available for each and every student for online support[3]. The Virtual Tutor was requested and will be used by "*Universidade Aberta*", a Portuguese e-learning course university, they want the Virtual Tutor as a tool for their students to use as a complement for their studies.

With the development of a prototype of the concept and the expected behavior we wanted the validity of our concept and idea as well as establishing the foundations of the empathic agent and the surrounding tool. We also wanted to find out whether or not this idea would have meaningful impact in our potential user's lives. We also investigate whether the relationship and rapport building aspects as well as empathy had an impact on the agent's reception by its potential users in order to find out if these factors were relevant to keep the users interacting more with the Virtual Tutor.

1.1.1 Problem

On a more generic note we believe that there needs to be a better always on support for students that are not adapting perfectly to university. On a more specific case, how will we mediate the interaction between student and virtual tutors so that the approach is successful.

1.1.2 Hypothesis

Our hypothesis is that this interaction is based on three aspects: relationship building, building rapport and generating empathy. On this work we explore a set of interaction scenarios based on these aspects with two virtual tutoring agents with complementary personalities that accompany the student continuously without forgetting to call the real tutor when necessary and test the reaction of potential users on what concerns the relevance of the expressed behaviors and the acceptance of the tutor's characteristics to examine if one of these characteristics is more important than the others for this particular context.

1.2 Original Contributions

With the Virtual Tutor we plan to contribute by:

- making a state of the art gathering on what concerns virtual tutors in the context of Rapport, Empathy and relationship building;
- developing an interaction metaphor with two Virtual Tutors with complimentary personalities and their usability tests;
- developing interactive scenarios that demonstrate the various ways that the interaction dimensions can be relevant;
- evaluate the relevance of the studied dimensions with potential users.

1.3 Thesis Outline

In the rest of the sections you will find the most relevant steps and decisions that allowed us to build the prototype, and correspondent usability tests for the *Virtual Tutor* app that was a foundation for the testing of the components of simulated relationship, rapport and empathy. Then you will find the experimental protocol, results and conclusion of our experiment to uncover which of the six chosen behaviors correspondent to the three main topics were more important for the users. From the results of this experiment you will find statistical tests and the conclusions taken from this experiment.

2

State of the art and Related Work

Contents

2.1 Virtual Tutoring Agents	xx
2.2 Relationships	xxi
2.3 Rapport	xxiv
2.4 Empathy	xxvii

This chapter will not only present the theoretical background but also other works on the field and state of the art projects and technology that inspired us to come to the prototype described in this work as the Virtual Tutor and the most important behaviors an agent for that context should have. All of these knowledge pieces carry a significance in some aspect or another that was key to our solution and tests. These related works and theoretical backgrounds are divided in four different topics that are the main topics that characterize our solution and testing areas: Virtual Tutoring Agents, Relationships, Rapport and Empathy.

2.1 Virtual Tutoring Agents

For educational purposes there are a great number of ways that we can apply agents: as personal assistants [4], user guides [5], alternative help systems [6], dynamic distributed system architectures [7], human-system mediators [8] and others. With the changes in the educational system, there is an emergence of dynamic and more complex infrastructures which need to be managed efficiently.

[9] proposed a paper in which they would develop a framework for a reliable and intelligent software that should have served a kind of tutoring function within a virtual team, their research's objective was to attribute an agent real human characteristics and behavior which in turn would enhance the agent's believability as a tutor and as a conversational agent alike. For that effect the agent would have the duties of a human tutor within a virtual team, these duties are:

1. Interrogator - Asking questions to the members of the group so they can answer them. These questions should help the members of the group in reaching a common goal.
2. Reviewer - Analyzing the answers provided by the team's members in whether these answers are correct or not. If any disagreement between the team's members occurs, the agent advises them to communicate in order to reach an agreement.
3. Monitor - Keeps a record of the answers from all members and of their communication with each other.
4. Instructor - giving individual instructions and helping the members who can not keep up with the progress of their group-mates.

[9] also argues that to corroborate the aforementioned changes in the educational system, new types of educational mechanisms and services need to be developed and supplied. These services need to fulfill the following requirements: customization, adaption and support to the users when they deal with new technologies. [10] argues that agents have emerged to provide better solutions to these requirements as compared with already existing technology and they can also influence different aspects in educational systems: they supply new educational paradigms, support theories and can be very helpful for both teachers and learners. [9]'s work is a non-hierarchical environment in the sense that everyone can teach and everyone learns. According to [11] in the beginning the main trend was to build an agent to replace a teacher in all the ways a teacher teaches. Now the

environments are more exploratory and that gives the students more responsibility for their learning process. [12] believes that the roles of an agent of this type must be to monitor, control and catalyze social knowledge building among the community of learning.

In [9]'s proposal every agent has its mental model which contains: emotions, personality, attitudes, objectives and knowledge. The proposed framework has plans, questions, topics and actions.

A plan is a course of actions in order to achieve one of the agent's objectives.

An action can either be primitive or complex. Complex actions are called decisions.

Each question has a set of possible answers and to each possible answer the agent will give a reply. When a question is asked, the agent will look for patterns of words and compare it with topics in its own base of knowledge so he can give the proper answer.

Each Topic can have subtopics. The learning material for a semester is a set of topics per lecture and the number of lectures. Topics may have precedence (which means a topic *b* may only be available after teaching a topic *a*). Each topic has an allocated time for it in lectures, associated questions, goals, results and consequences.

In our proposed virtual tutor we will be inclined also to provide a form of coaching and stabilization of the emotional status to our students.

Coaching is defined on [13] and [14] as a form of development between a coach and a client or coachee. The coach helps the client achieve professional or personal growth through training, advice and guidance. Most of the times the coach does not train the client in the specific subject they are accompanying, instead they take care of the client's state of mind and guide them to harness their potential and making the most of it.

The virtual tutor will use a set of questions and pre-programmed answers to keep up with the student's academic progress. It will analyze and monitor their answers with the option to warn the human tutor responsible for the student if needed, and will also instruct the student in academic and non-academic levels (i.e. it can tell students to study but can also encourage them). We agree that a virtual tutor should not replace a human tutor or teacher, the virtual tutor will be a coach and a guide to the student while also shedding some lights on academic matters.

2.2 Relationships

In order to make a virtual companion for a human, we have to find out the factors behind the forming of a companionship bond between humans so we can simulate it in the agent.

According to [15] a relationship is generally defined as "a persistent construct, incrementally built and maintained over a series of interactions that can potentially span a lifetime". [16] refers that these interactions can be placed in two groups: generalized patterns for an interaction based on the stereotypes of society (doing something that society believes you are supposed to do for a relationship) or interactions specific to that two-people relationship (some friends have specific interactions that they do not have with anyone else). Because of this last type of relationship it is difficult to say when a relationship starts or ends or even what defines the relationship between those two people. For in-

stance, two friends may talk rudely to each other and insult themselves and laugh it off, for some this can be appalling but for the integrants in that relationship is a perfectly normal behavior.

Since we are trying to prove that a motivated and happy user will interact with our virtual tutor more often, it is relevant to look at the way relationships between humans are formed because according to [17] relationships are a relevant aspect for "human happiness, physical and mental health", in [16] we also see that some of the benefits of a relationship for a human being are that relationships can provide emotional and instrumental support and other forms of nurture.

Levinger [18] proposes that human-human relationships go through five steps in that relationship's lifetime: 1- acquaintance, 2- buildup, 3- continuation, 4- deterioration, 5- termination. Since we are going to be focused on the creation and continuation of a good relationship with a human being, we will need to focus on the first three stages that Levinger proposed.

Starting with acquaintance, the process of meeting someone for the first time, it is common sense that the first impressions will dictate our willingness to get along with the person we have just met. That is why most people already know that first impressions are the most important aspect of starting to build a relationship with someone. If the first impressions are good, we are more willing to get to know and even agree with that person, however if our first impressions are not that good, we will be less willing to have any kind of relationship with that person.

As for the second and third stages proposed by Levinger, they have to do with the maintenance of a relationship. We maintain a relationship by ensuring we feel satisfied by its existence.

Many researchers state that we have two different types of behaviors when trying to maintain a relationship, routine behaviors and strategic behaviors[19]. [15] states that routine behaviors "serve to maintain the relationship as a side effect" (e.g. you and a friend see each other because you both go shopping at the same supermarket so you decide to go together). According to [20], strategic behaviors are used with the sole purpose of either maintaining or improving the current relationship, such as talking about said relationship or recalling happy memories. In our case we planned to measure the frequency of the contact between the student and the Virtual Tutors as one of the success rates of the application. We also implemented some behaviors like recalling past memories or commenting on the time between each contact in order to motivate the more eager students to interact more with the tutor.

Although we don't need to focus on the deterioration and termination part we have to be aware of them in order to successfully counter them and delay as much as possible the end of a relationship that can be good for the user. In [21], Bickmore bases himself on the human-human relationship research to point out some strategies that agents can use in order to maintain a relationship with a human. These strategies include performing new activities with the human, meta-relational communication (i.e., talk about the relationship), empathy, reciprocal self-disclosure, the use of humor, talking about past and future events, continuity behaviors such as greetings, and emphasize commonalities and de-emphasize differences, to increase solidarity and rapport. On that same article, Bickmore states that empathy is really an important factor for close relationships.

In the context of games [22] proposed a model of agent's behavior to allow them to reason about

their own behavior by studying, among other things, the relationship between NPCs. In this model, the agents behavior and liking toward one other agent are influenced by the agent's emotions and affected by their personality. For instance, if one agent *B* causes a negative emotion on an agent *A*, then that would decrease the value of liking that agent *A* has for agent *B*. However, this model does not accurately translate how a relationship between an agent and a human is born and develops.

To study that exact factor, [1] shows us a framework for the creation of a realistic social game-agent companion called *ERiSA*, a framework that grants the agent the ability to establish simple social relations with a human throughout the course of their interactions together. This framework is built with five different components: the Sensing Component, the Interpreter Component, the Behavior Component, the Agent Components and the Game Components. The Sensing Component provides modules for the sensory inputs received from the user. In *ERiSA*'s case the sensory inputs are face recognition, facial expression recognition and speech recognition. This Face Recognition module tells the users apart making the agent recognize each user by their faces. The Facial Expression Recognition is based on FACS and six basic emotions [23].

The Interpreter Component receives data about the player: their facial expressions, utterances and identity. Their utterances are interpreted by a Verbal Interpreter Module while their facial expressions are interpreted by a Non-Verbal Interpreter Module.

The Behavior Component defines the agent's behavior based on the information of both aforementioned interpreters and the agent's Interaction Rules. It is then sent to a SAIBA compliant embodied system [24] by the Action Selector Module. After that the Action Manager updates all necessary modules.

The Agent Component is divided in memories of internal rules, states and personality and a Social Relationship Module that manages the relationship between the two players at the time whether they are human or agents. Finally, the Game Components are everything necessary to have knowledge of the game such as rules and game states.

[1]'s agent has a personality based on the OCEAN personality model [25]. Although there were some limitations to this experience, they found out that an agent using this particular framework was able to establish a relationship with the users.

They developed a formula to implement a social relationship based on emotion, likes and personality traits. However we want the people to feel attached to the virtual tutor therefore we believe it is better not to use such a model because it could lead to aspects where the virtual tutor becomes less empathic due to being mistreated by the user. Using these studies on relationships we will try to make the agent create a relationship with the user in order to make the user want more interactions with the agent. With this we hope that the agent will improve the users emotional and affective stages by making so that the agent can become the user's friend and they can rely on it for their good and bad moments. To increase the relationship, the Virtual Tutor will keep a memory of former interactions and call the user by their name. It will also make some self-divulging comments to increase fondness and we believe that as time and interactions happen it will become friends with the user.

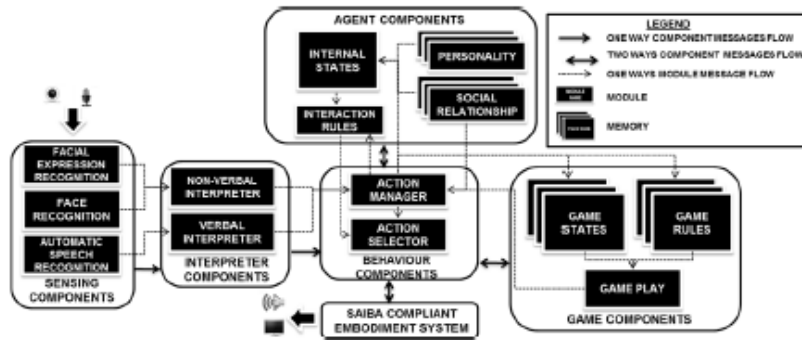


Figure 2.1: [1]'s ERISA's framework overview

2.3 Rapport

The feeling of harmony, sync and flow that humans feel when they are engaged in a good conversation with someone is what we know as rapport [26]. Also according to [26], rapport mechanisms are what keeps the interaction interesting for both parts.

In [27]'s study about Rapport, they establish that in order to create this feeling of Rapport three factors must be considered: positivity, mutual attentiveness and coordination. If positivity is a strange word you should consider it as the capacity of an agent to be friendly towards the user. Only when these three aspects are working together can we build rapport. Tickle-Dengen and Rosenthal[28] mention that some of these rapport mechanisms include smiles, head nods, mutual attentiveness and coordination (for example mimicking posture or using the same gestures). They also mention the three aspects of rapport on study [27], describing positivity as a feeling of approval and friendliness, these feelings can be achieved through humor or self-disclosure. They describe mutual attentiveness as the feeling we get when the other is solely focused on what we are saying. And coordination is described as the feeling of sync and predictability (such feelings can be induced by mimicking posture or the respect of social stereotypes). Studies [29], [30], [31], [32] and [33] each show a different area of social interaction that rapport facilitates: negotiations, management, psychotherapy, teaching and care-giving. Studies [34], [35] and [36] show that agents with rapport are more often perceived as more engaging, promoting of fluent speech[34, 35], according to [37] these agents are found to be more persuasive and according to [38] reduce user's frustration while interacting with said agent. However, using only one of the three aspects can backfire and cause a negative social effect[27], this means that we need to use the three aspects at the same time to build rapport if we want to achieve a positive social effect.

According to [26], among human dyads, rapport can be conceptualized on happening on three levels: emotional, behavioral and cognitive.

They explain that emotionally, rapport is a rewarding experience due to the flow and harmony we experience. That cognitively there is an understanding with our conversation partner. That behaviorally there is a coordination of movements with our partner.

When starting any kind of relationship with someone, positivity is a massive decisive factor for

building rapport and carries a huge importance because both interactors are behaving the way society expects them to: greet and be polite. However, as time and the relationship progress, positivity will lose its importance because both interactors expect less friendliness from each other than on the first meeting [28, 39]. [39] also states that as time moves on and the relationship becomes stronger, both interactors may lose what we see as supposed behavior (or society imposed stereotypes) and form their own way of interacting. We can see it clearly if we think about those friends that treat each other with sarcasm and/or insults.

Eye-contact and gaze is one of the ways to build this feeling of rapport. [40] made an experience in order to see the actual effects of gaze and eye-contact on people by using a mutual gaze agent that synchronizes gaze behavior with pre-recorded voice and gestures. They concluded that participants would recall a story better if the agent looked at them more often. But also women felt better when the agent gazed less at them.

According to [39], in a context of tutoring, if the tutor is a friend of the student, the student will prefer if the tutor is a bit rude and more informal. However if the student and the tutor have no prior friendship whatsoever the student will prefer a more traditional, conservative approach. We plan on making the tutor behave more familiar with the student as the number of interactions grows, in order to emulate this information. Meaning that the more familiar the tutor becomes to the student, the more familiar and informal its speech will be. We plan to take this method into our Virtual Tutor so that as the number of interactions grow, the Virtual Tutor takes a more informal even "rude" approach with the user.

[2] created a Robot with a Rapport module that followed the three main aspects of building rapport. The agent should adapt to the surrounding society, behaving the way society expects it to. Their rapport module has the following components:

- **Dyadic State:** registers the partner and environment's information.
- **Perceptions:** Perceptual Information
- **Rapport Effectors:** components that generate signs of rapport proposing actions to the *Rapport Controller* given the dyadic state of the interaction.
- **Rapport Controller:** manages the rapport actions given by the *Perceptions*

This controller manages a quintuple $A = \langle G, P, E, I, T \rangle$. G stands for **Group** and it specifies which part of the agent's body it's going to manipulate. They mention that for example a *Speech Group* action such as speaking should not interrupt a *Face Group* action such as a facial expression. P stands for Priority and it corresponds to a relative priority for each action proposal. E stands for execution description and it describes how the agent will execute the proposed action. I stands for Interruption Description and it describes how the agent will interrupt the proposed action. And finally T stands for Timeout and it concerns the maximum time value of the proposed action. Periodically, the controller takes a snapshot of the ongoing and pending actions received by the different effectors. At the time of the snapshot or when the agent receives a new action proposal, the controller must



Figure 2.2: [2]'s Goal Tree of the Rapport Model

analyze which actions should be interrupted and replaced. If there is a conflict (i.e. different actions on the same group), the action with the lowest priority value will be interrupted. Also, if an action takes more time than it should (regarding the Timeout value), it should be interrupted. In order to enhance mutual attentiveness their agent would produce listener behavior and make eye contact with the listener. Their *Mutual-Gaze Effector* is based on the works of [41]. It swaps establishing eye-contact with the participant and the game during established periods of time depending on whether a task is being performed or not and on whether the user is an introvert or an extrovert. They also use a *Backchannel Effector* based on the works of [42] to analyze pitch variations during interaction to produce listener behavior.

To enhance positivity the agent will make context-related vocalizations to motivate or praise the interaction partner. These utterances can also be of the self-disclosure type as suggested by [43] that self-disclosure has an important role in closing relationships between two strangers. And to enhance the final aspect of rapport, coordination, the model is composed by three different parts: behavioral mimicry, basic conformance to social norms and the already explained backchannel. The model mimics head gestures and facial expressions, these animations are proportional to the perceived emotional intensity. [40]'s agent adheres to these following social norms:

- Introduces itself when it meets someone for the first time.
- Greets the interactor before starting the interaction.
- Avoids invading someone's personal space.
- Says "please" when making a request.
- Says "thank you" when a person does a task asked by it.

This work showed no statistical proof that the applied rapport strategies increased likability. However they postulate that if they could have increased the sample size, they might have obtained a clear confirmation of said improvement on likability. However when they analyzed the video feed of the interactions with the users they noticed more frequent positive reactions from the participants in the rapport condition.

With the addition of rapport to our virtual tutor we plan on making it more human-like, in order to gain the user's trust and also more engaging to talk to. We believe that if the user is more trusting of the virtual tutor and is able to speak freely, the user will be more motivated to continue interacting with the tutor and, if this project is successful, the user will also have an improvement in his or her school-life. In order to increase the positivity factor of rapport we used the strategy of self-disclosure,

in which the agents commented on their own emotions about the given context. In order to generate the feeling of coordination and mutual attention components of rapport we decided to implement a back channeling mechanism on the utterances of the agents, to seem like they were reacting to the user's input and to each other's sentences.

2.4 Empathy

One of the most important components of the virtual tutor will be an empathy behavioral component, we give a major importance to empathy because [44] mentions that empathy is argued to facilitate the development and creation of social relationships. Since we want the user to establish a good social relationship with our virtual tutor, we believe empathy is key on achieving this.

According to [45], empathy increases fondness, similarity and affiliation. That is why it facilitates the development and creation of social relationships. Empathy also plays a crucial role in cooperative and prosocial behaviors such as helping and caring [46].

[46] gives us a broad definition of empathy as "an affective response more appropriate to someone else's situation than to one's own" and [47] adds that not only affective processes are influenced, but also cognitive and prosocial behaviors. An example of the former is for example how we take actions to reduce the other's worries or distress. Work [48] showed that empathic agents are often perceived as more likable, caring and trustworthy than non-empathic agents. Study [49] also shows that empathic agents can foster empathic feelings on their users. In [50] Paiva et al. shows us that empathic virtual agents can stimulate empathy in children helping them deal with bullying situations.

Hoffman [46] also states that empathy results from five different kinds of behavior: motor mimicry, classical conditioning, direct association of cues from the person we are empathizing with, mediated association and perspective-taking. Suggesting that from these five, perspective taking is the most empathy arousing one.

Empathy has been considered one of the most important determinants of pro-social behavior [9]. Prosocial behavior can be defined as "voluntary behaviour intended to benefit another, such as helping, donating, sharing and comforting" [4]. We need our agent to have pro-social behavior in order to help coaching its users. We also plan on using empathy to create such pro-social behaviors in our virtual tutor, in a way it can motivate and help the users keep a high morale and affection level.

[19] conducted an experiment in which they made a robot with an empathic artificial intelligence to study long term interactions with empathic robots. They evaluated this agent through a series of chess games and chess puzzles with children. This robot's A.I had a chess heuristic, which means it could look at the game state and find out who had the most advantage and what the best move normally was. By doing this taking into account the user's point of view (putting itself in the user's state). Because of this and some visual recognition, the agent was able to infer the user's affective state and act accordingly. The point of the robot was not to be a competitor but to be a friend to the children by cheering them on and giving them support on their plays or indecisive times. Some children said the iCat (the robot in question) was teaching them to play and at the end of the experiment some



Figure 2.3: I. Leite's iCat interacting

children even considered the iCat a friend. There were many factors that made the iCat achieve these results, but one of the major factors that distinguished the iCat was the empathy component. With and without the empathy component the results were different and the empathic iCat was perceived as being more friendly and more trustworthy. Most of the children that interacted with the iCat wanted another interaction. These results are part of why we think having an empathic virtual tutor will keep the users engaged throughout their interaction cycles. [50] built an empirical computational model of empathy which allows an artificial agent to exhibit different degrees of empathy. That model is divided in three processing steps: the *Empathy Mechanism*, the *Empathy Modulator* and the *Expression of Empathy*.

1. The *Empathy Mechanism* produces an empathic emotion.
2. The *Empathy Modulator* modulates that empathic emotion.
3. The *Expression of Empathy* triggers the agent's behavior through the modulated empathic emotion.

[50] makes the evaluation of this module was made on the MAX [51] and EMMA[52] context with a human partner. The evaluated situation is as follows: EMMA has the empathy module and the human user and both agents talk to each other. On this occasion the user can cause either positive or negative feelings on both MAX and EMMA by saying positive or negative utterances towards them. EMMA will observe MAX's facial expressions and will emphasize with the emotions he is displaying. This experience had the objective of evaluating how the agent's behavior produced by the empathy model was perceived by the human user.

As mentioned before, their empathy module was divided in two parts. The *Empathy Mechanism* consisted of "an internal imitation of perceived facial expressions" and "resulted in an emotional feed-

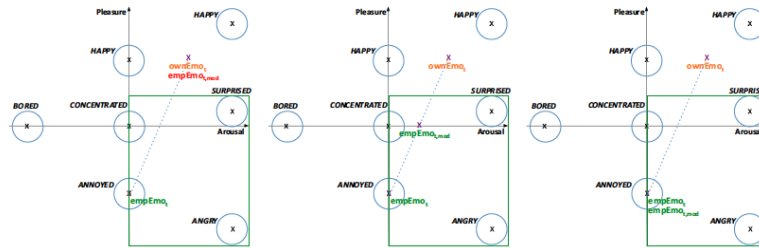


Figure 2.4: Maxx and Emma's interaction

back that represents the perceived emotional state", which means EMMA will analyze MAX's facial expression and take it into account when modulating her own empathic emotion. On the *Empathy Modulator*, EMMA takes into account MAX's emotion as well as her own to modulate a new empathic emotion towards what the user has said, this emotion will then be expressed by the *Expression of Empathy* with an utterance, breathing and blinking. The higher the arousal EMMA feels, the higher the frequencies of EMMA's blinking and breathing. This arousal would be higher depending on the degree of liking EMMA had for MAX. Three different values of likability from EMMA towards MAX were tested: EMMA had a likability value of 0 towards MAX, EMMA had a likability value of 0.5 towards MAX and EMMA had a likability value of 1 towards Max. In the scenario in which EMMA had a higher likable towards MAX, EMMA's emotion was more evident in comparison with any lower values of likability towards MAX. This EMMA is also perceived as more empathic towards MAX. The users identified different values of intensity on what concerns EMMA's empathy towards MAX. However the EMMA who was deemed more empathic towards MAX was only partially the more likable one of all the three tested modules.

A state recognition and consequent perspective taking is one of the most crucial aspects of our work. With it we can infer from the user's grades, upcoming projects and past interactions whether the user's emotional state is positive or negative. As mentioned already, an empathic agent is normally perceived to be more likable by the users, we plan to use this likability to improve the frequency that the student uses the proposed virtual agent tool and as a result perhaps improving the user's results in their studies.

3

The Virtual Tutor app

Contents

3.1 Approaching the Concept	xxxii
3.2 Usability User Tests	xxxvii
3.3 The final Prototype	xl

3.1 Approaching the Concept

The stoppage of contact between university students and their assigned tutor is indeed a frustrating motivational problem for some students. In order to mitigate this frustrating problem we started building a Virtual Tutor with coaching skills with the objective to take emphasis on three fronts: empathy, rapport and building a relationship with the user. We believe that by focusing on these three fronts our Virtual Tutor will provide emotional support to the user and make them more inclined to continue interacting with their tutors and as a result helping the users with their daily academic challenges.

3.1.1 The Quadrant System

Our agent would have access to their assigned student's most relevant academic information most likely with a future integration with an online platform: test dates, test review dates, student's grades and the number of courses the student is taking that semester as well as the course's level of difficulty, expected effort for the student and the importance of the course in itself, the importance of the course will establish a priority between the various strategies for the courses. The virtual tutor would also have a log of the previous interactions with that user, so he can start "profiling" the student's personality and finding out what helps the student the most. On the first interaction the virtual tutor would set "How easy do you think this course will be?" and "How much do you like this course?". This way, for each course the student would be gauged in two parameters "*confidence*" and "*motivation*" respectively. With these two values the Virtual Tutor would set the user in one of four possible quadrants: OK, Motivation, confidence and Both, as you can see on figure 3.1.

If the student's both *confidence* and *motivation* values for that course are positive, we say that the student is in quadrant OK and the Virtual Tutor only has to make sure to maintain the student's suggested working hours and motivation. If the student's *motivation* value is low but their *confidence* value is high, we say that the student is in quadrant Mot and the Virtual Tutor has to increase the empathy and encouragement components while maintaining the suggested working hours given. If the student's *motivation* value is high but their *confidence* value is low, we say that the student is in quadrant Conf and the Virtual Tutor has to increase the suggested working hours given to the student while maintaining the empathy and encouragement levels. Finally, if the student's both *motivation* and *confidence* levels are low, we say that the student is in quadrant Both and the Virtual Tutor has to increase both encouragement and suggested working hours levels. The agent will then assume that a lower grade at one course that the student is on quadrant OK is more serious than a lower yet still passing grade in a course in which the student is on quadrants Mot or Conf and a lower yet still passing grade in a course in which the student is in one of those quadrants is more serious than a lower yet passing grade in a course in which the student is in quadrant Both. The courses are initially set by the student on the four different quadrants, however, the Virtual Tutor can change the quadrant it believes each course is for its student. These changes can be influenced by the results on these course's grades (if a student stops studying and gets bad results on a course he was on Quadrant Conf, the Virtual Tutor will assume that the student's motivation has lowered and therefore lower the

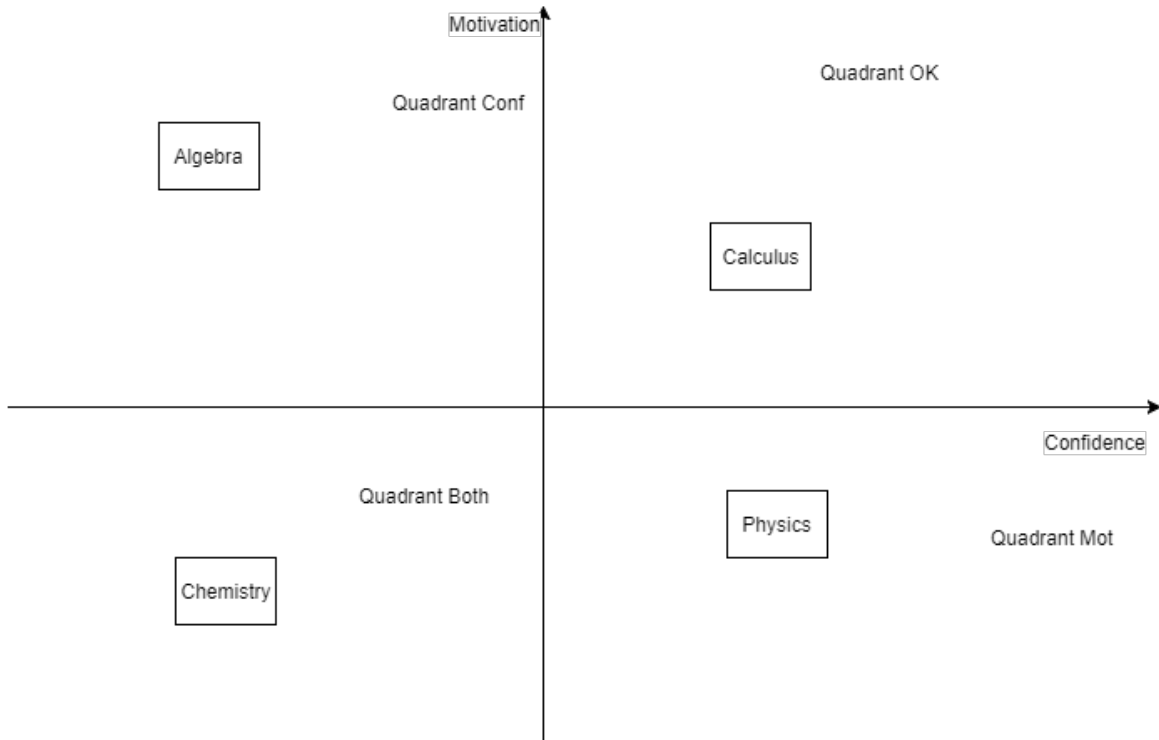


Figure 3.1: Example of the four quadrants for student's motivation and confidence

value of the variable concerning Motivation). These variations can cause a change in the quadrant for that student for that specific course.

3.1.2 Impacts of confidence, motivation and importance

These parameters of *motivation*, *confidence* and the course's *importance* alongside the tutor's personality and the user's grades history with that course will determine an expectation for the tutors on the next grade for the next evaluation the student would take. The different distances between the agent's expectation and the user's actual grade will trigger different reactions on the Virtual Tutor's part to the input or consultation of the new grade. The agent's reactions will also be higher or lower depending on how permissive the agent is towards that course, i.e. how much of a priority the course is.

During any interaction the student can change the values of *confidence*, *motivation* of a course. This feature was decided because not all courses are what we expect them to be in the beginning of the semester and also because the agents may not have the right estimate all the time, so it is up to the student to be honest with himself and the agents and input the right estimates of their *confidence*. However, the course's *importance* cannot be changed by the user and is defined by the professor responsible for the course. This decision was made as to not make students put all courses as unimportant only to get a less strict response from their Virtual Tutors.

3.1.3 Normal scenario of interaction

Every interaction, if the user did not come to talk to the Virtual Tutor from a pre-established arrangement, after greetings and formalities, the agent would gather from other time between the last interaction and this interaction went (of course studywise), how much study the student got done, what courses did the student study for. In the case of an event (i.e. after taking a test, after test reviews, after the student receives a new grade) the agent will ask about it, try to ascertain the student's emotional state after that event and ask the student (so he can admit) what went wrong (if anything). When the student gets the grade corresponding that test, the Virtual Tutor will know and it will try to comment with the student about it to generate empathic responses and establish a sound plan for now on. Some time before the test the Virtual Tutor will remind the student about this test and start asking how much time the student spent on studying for that specific course and the app itself will send notifications to the user about that test. All these interactions are made to be more engaging, realistic and to leave the student comfortable around the virtual tutor by using a rapport component.

Using the grade values and the course's settings (motivation, confidence and importance), the empathy component would make the virtual tutor take the student's perspective by taking into account the student's grades and their progression as their emotional state. For instance, if there is a big decrease in the student's grades the virtual tutor will take it as a decrease in the student's emotional state, on the other hand if there is an increase in the student's grades the virtual tutor will take it as an increase in the student's emotional state. By empathizing with the student's state, the Virtual Tutor will "feel the same pain" as the student (i.e. it would treat the student's failures as its own). The Virtual Tutor would always try to maximize the student's emotional level via cheering them up in just the right amount of encouragement and reminding them just enough. If it gets to a point where the virtual tutor gauges that the student's emotional status is low and nothing it can do seems to help, the virtual tutor would contact the student's assigned real tutor so he can intervene and talk directly to the student.

3.1.4 Managing Emotions

In order to give the agent characteristics that the students can empathize with, to make the tutor less program-like and also to better empathize with the students and generate Rapport with them, the Virtual Tutor will possess its own Affective Model. This Affective Model would consist of its emotional level that he also tries to maximize. This emotional level would be influenced by the Empathy Module (for instance if the student gets a bad result, the student will feel sad, therefore the Virtual Tutor will also feel sad). The Virtual Tutor's emotional state could be told to the student by the Virtual Tutor's self-disclosure utterances in order to create Rapport and increase the relationship between the two. The Virtual Tutor will care for its students, as such its objectives are that the student's emotional status is high and that the student is academically successful. By building a relationship with the user we want them to care for the tutor as the tutor cares for them and we want that to be a motivation factor to keep interacting with the Virtual Tutor.

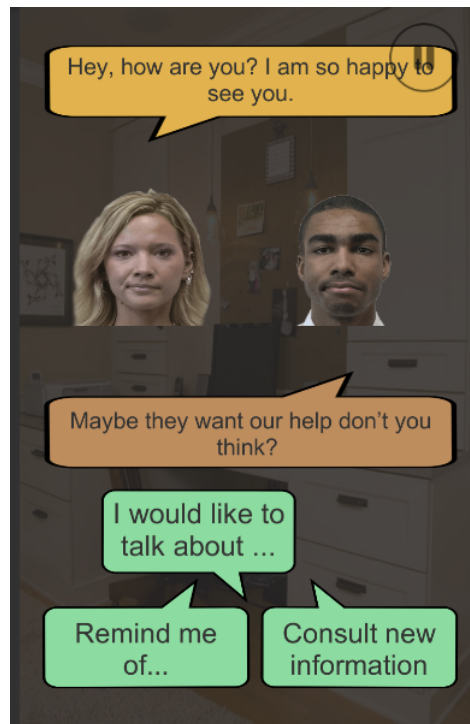


Figure 3.2: The two tutor agents in our prototype

3.1.5 Design and Interactivity

On a perspective more on the Design level we had several concerns about the interactivity and enjoyment of the interaction loop between the user and a single tutor. These concerns were mostly detected on two levels: the Virtual Tutor itself and the way that the Virtual Tutor interacted with the user.

The main issue with the Virtual Tutor in itself was that even if it were the most dynamic and engaging agent in the way it held a conversation we believed there was more potential to be explored on what the interaction with a virtual agent is concerned. The limitation of it only being one virtual agent was that it narrowed down immensely the ways the agent would have to communicate with the user as for if there is only one agent and one user the communication would have to be direct and the every remark or question we had to take into consideration that the Tutor would be talking to the student. As such, we felt extremely limited with only one virtual tutor and on several meetings we posed the idea of actually having two different tutors with each their own personality and set of reactions and manners of speech. We finally reached the consensus that one of the Virtual Tutors should be more permissive, more forgiving, kinder and gentler as opposed to the other tutor that would be more exigent, a little more sarcastic, and less permissive as the other tutor. This idea made us be more creative on what the interactions could be like because we could make different types of contact by making both tutors talk to each other sometimes instead of with the user.



Figure 3.3: Example of the dialogue in Oxenfree

3.1.6 Dialogue

However, even with these changes we still had the idea that the interaction between any of the tutors and the user would feel a tad stale, and after careful consideration we realized why: the dialogue in itself was reactive and only functioned in a way that the agent and then waited for an answer, and then reply accordingly. There was a lot of downtime in this interaction and we feared that the users could get bored or tired of interacting with their tutors. As a result, we tried to improve the dialogue system so that it would not cause this effect on the users so we decided to make the dialogue system run like it does in the video game Oxenfree. Oxenfree is a 2016 supernatural mystery graphic adventure video game. This game's dialogue system has the dialogue happening during gameplay instead of in separated cutscenes.

During the time the dialogue is going, the user's character has some input options, designed as speech bubbles of different colors, above his head and can decide whether to chose or not one of them. If the user decides to chose one of the options, that will be the main character's line and that will have the game's defined consequences. However, if the user does not choose any of the choices, the dialogue will just continue and the characters will just assume the main character has nothing to say. During the other character's interactions more input options will appear over the main character's head as possible comments or answers to the subject in question. We inspired ourselves in this system to make our own. The agents will normally talk to each other about the topic at hand and they will pose one or two questions to the user, the user will have options available at the beginning of that interaction and if they have already chosen their desired input they can just click it without having to wait for the conversation to be over and the agents will take that answer as a priority and react. However, if the user does not choose an answer right away, the conversation between the agents will continue and some new questions may be asked to the user which will cause more options to appear. Finally, if the user does not respond for a fixed period of time the agents will have two different reactions depending on which *Topic* is being discussed: they can either assume the user does not want to answer the

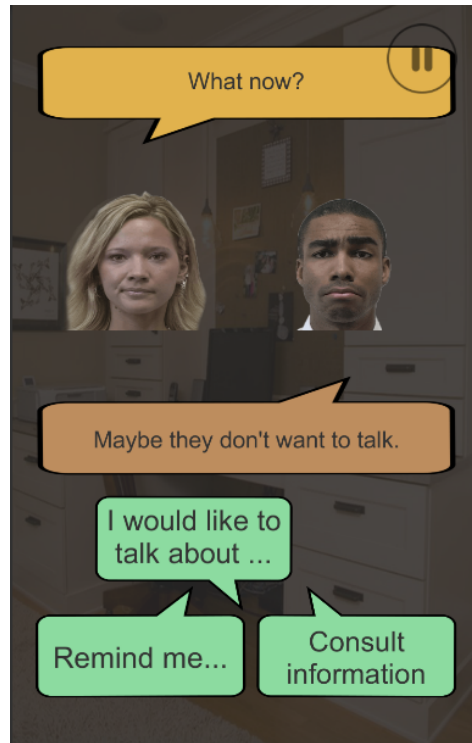


Figure 3.4: Example of the timeout scenario in the dialogue

question and will understand it, mention it in order to build empathy and show respect towards the user's choices, and then change the subject, or they can assume the user is not paying attention to them and after a few comments about it they would close the app.

3.2 Usability User Tests

3.2.1 Presenting the first functional prototype

As the Virtual Tutor app's target audience was University students facing a semester in University. As such, we believed it would be in our best interest to get feedback from people behind a few Universities in Portugal while attending a meeting of their group. In this meeting we presented our first functional prototype and took notes from feedback provided by the other teams, especially the one that requested the "Virtual Tutor" agent.

The prototype was received with mixed feelings, especially of surprise because of the liberties we took that were not what the committee had imagined or was expecting would happen, especially the fact that we had two agents instead of the expected one agent and the fact that the sprites were based on cartoons for this prototype. In general they were surprised and wary of some choices we had made. However, they were also pleased with the attention we gave to the student's emotional state and commented on the usefulness of the prototype.

The main comments that were mentioned were:

- The way the courses were structured differed from the implemented way and we were missing

some key elements.

- The requirement was an agent that was also capable of teaching, aside from the emotional management component.
- The dialogue was meant for people with an average age way lower than their students' average age.

In what concerns the course structure, we have been told that there are other relevant moments of academic importance throughout the semester that were not evaluation moments, such as every relevant achievement decided by the respective course's responsible professor. On the second version of our prototype we would attend to this issue and fix it in a way that was pleasant to the committee.

Some of the people that spoke were expecting a Virtual Tutor that had a pedagogic component implemented on it or a module that would allow it to teach or answer questions related to the subjects taught on each course. However, we decided to focus more on a coaching side that was deemed as important and was more related to this document's objectives. As we were writing the script for the dialogue between the agents, we took the liberty of adding sarcasm and humor on the retorts of the agents towards each other and a little bit of less positivity on the more exigent agent's personality so he would almost sound like a grouchy old man that was always making fun of the more bright and optimistic agent's point of view. The agents were viewed as juvenile and too immature for the average age of the students in the main university of the committee in question. We agreed with this after reading more about statistics from their students and on the second version of the prototype we decided to change the script to be more mature and formal in order to not break the interaction and positivity with more mature users.

3.2.2 The MoJo

On May 30th, we participated in MoJo (Montra de Jogos) in Instituto Superior Técnico - Taguspark to test our functional prototype with users for the first time (see figure 3.5). We managed to test the prototype with a total of 15 people, 13 of those were male students with ages from 18 to 25, one of those was a female university student, one of those was an artificial intelligence professor at Instituto Superior Técnico and the other one was a CEO of a Portuguese video games company. The testing took place between 10 a.m. and 5 p.m. and were done in an open area with people that took interest in the prototype.

3.2.3 Setting up the experiment

Above all else, we were looking for concept validity from potential users and any experience issues or technical issues with the interfaces. In order to determine what we wanted to know, we decided to adopt a method of open interview based on the chapter related to testing of the book [53]. To set up the interview we decided that one of us would be the interviewer and the other one would be the analyst. The interviewer would connect a web cam filming the user's reactions while testing the prototype and accompany the users through the interaction and asking open ended questions



Figure 3.5: Mojo Participation

Comment	Times mentioned
Change input balloons design	6
Agents' feedback is good and not too proactive (their job)	6
Noticed emotions	6
Useful Idea and concept	5
Integration is good	5
UI is intuitive	4
Agent design is weird	3
Did not notice emotions	3
Too proactive	3
UI is not intuitive enough	3
Agents could disappear when they are not talking	2
The agents were helpful	2
Liked the sliders	2
Gibberish is bad	2

Figure 3.6: The experiment's results

employing some strategies mentioned in the same book in order to receive answers in the intended testing topics. The analyst would be in a room recording the user's reactions and answers for further testing. Unfortunately the MoJo's conditions were not favorable to this kind of testing due to low Internet connection and much background noise so the interviewer would switch during the day and he would take notes during the interview with each user. Each user would be given a context page to read (see appendix 1) and then 3 tasks to perform to see how well they could use the application. These tasks involved: seeing the grades they already got without going to each course menu, adding a new evaluation result for any course and setting up a study plan for any course.

3.2.4 Results

From this batch of tests we came to some conclusions about our prototype, firstly this solution we tried to implement is conceptually not for everyone: some students will feel at ease with the coaches and feel empathy towards them as others will be intimidated by the proactivity and say it is too much for an application to do. The main reported problem had to do with the input dialogue bubbles of the prototype, their colors and shape. Since each option was also a speech bubble like the ones from the agent and had a different color and it was different from any of the agent's lines color some users did not understand that these were actual input options and stared at them wondering what to do. This result also has to do with the fact that some animations were not yet implemented and some timings

were a bit off. We also received the concept validation we expected, as most people commented on how useful the concept was and how much students in general could use it. Another thing that was generally praised was the integration with the Android ecosystem especially the integration with Google's calendar, people were impressed and commented it is easier to use if it relies on apps they use every day because they are already familiar with them. For the tested prototype we decided to express more emotions without just recurring to the agents changing expressions. To that end we decided to use a simlish module we had downloaded. For that reason we decided to implement sound and add them a simlish sound mode. Simlish is the popular gibberish language spoken by the characters of the video game franchise "*The Sims*", we found a fan made module of some sentences with two different voice pitches and we added each of the spoken sentences to either expression, so whenever there is a change in an agent's emotion, the agent will make a meaningless but emotive sound referent to their new emotional state.

After the MoJo tests we concluded that the simlish utterances were described as annoying for two reasons, the fact that they seemed like real words so the users would be frustrated trying to listen and not understand what was being said and also the fact that the sounds in themselves were screechy and repetitive. In order to fix this annoyance we decided to change it completely and adapt a TTS (Text to Speech) module.

Lastly we had several times the suggestion of making the application draw the student's performance graphs in each course throughout the semester or the year.

3.3 The final Prototype

After going through all the testing procedures mentioned on the subsection above and analyzing their conclusions we reached a final solution that took the concepts in the metaphor as goals and went off from it. Our Virtual Tutor app was built on Unity3d using scripts on the C# programming language and using Unity's user interface options. We chose Unity3d because it had compatibility with 2D programming as well as a component to develop the UI - user interface component of an app or game, it is also an engine that can port the prototype to mostly any platform and it works well with the necessary Android plugins, for example to implement the idea of the integration with the Google calendar for the study plans.

3.3.1 Dialogue

One of the key factors of the implementation on structuring dialogue was made by creating an object we called *Topic*, each *Topic* would be the highest object to dialogue in the hierarchy and would contain a set of *Lines* and a set of *Inputs*. This *Topic* served as a divider of the interaction in subjects. For example, when the Virtual Tutors greet the user, they are in the topic related to greetings, and in that topic there is information on the *Lines* they have to speak relevant to the greet the user and there is also information regarding the input choices the user has. Every time the user choses one of the input options the *Topic* is changed. Each *Line* object has information about which string of text needs



Figure 3.7: Example of Checkpoints in the UI

to be "said", which tutor needs to "say it" as well as the start and end timing of that line, like a movie script or a movie subtitle program. Because we need to know which tutor has to "say" which sentence we have an *Agent* object that knows to identify its own position to know which agent it is and its own emotion. Lines also have their own timing, we implemented this timing feature for two main reasons: timeouts and animations. Regarding timeouts it is easy to realize that if we want to keep the Oxenfree effect we would have to make the lines flowing and the conversation going without the need for it to be a user input. As such, in each topic there are more than one line for each tutor as they initiate a conversation amongst themselves whilst trying to catch the user's attention and posing the questions they want answered. If the user does not answer after a fixed time after their conversation has ended, the tutors depending on the topic at hand will do one of two things, assume the user is not paying attention, comment on that and close the app; or the tutors will assume that is a delicate topic that the user most likely does not want to talk about and change the subject. As this is a flow conversation instead of a Q&A machine, all lines need to have a right timing to appear on screen in order to fit the context of the moment and provide meaningful solution. As such, all lines and possible Inputs have a start time and a disappear time.

3.3.2 Checkpoints

We also created the *Checkpoint* object to encapsulate both the formative evaluations and the achievements deemed relevant by the course's responsible professor. We developed an abstract class called *Checkpoint* and these "checkpoints" would have an associated grade to it or they would just be items to check on a checklist. The evaluations correspond to pretty much all evaluation mo-

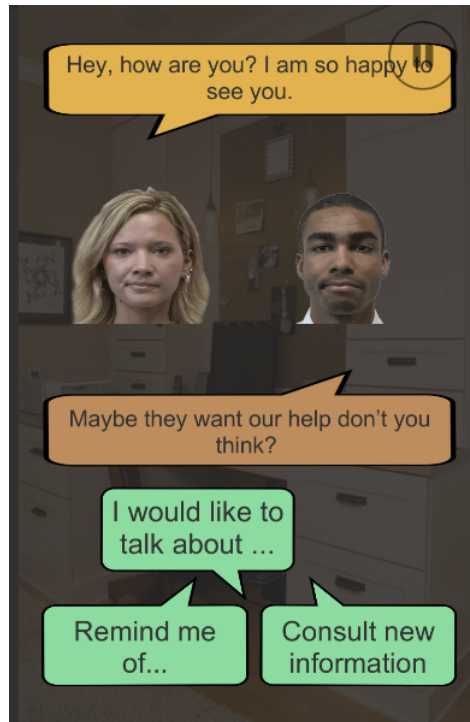


Figure 3.8: The final Avatars

ments, be it exams, tests, mini-tests, essays or even projects, i.e. everything that can have a score. On the other hand, the important events correspond to events that can have a state of either true or false, as items on a check list. Examples of these events can be downloading the research provided by the professor for a certain topic or attending the study session to post questions before a certain evaluation moment. Although these events do not have an associated grade with them, we felt it too important to be left out from the student's goals. Both of these events would be shown in the respective course menu under the "Checkpoints" subsection and be sorted by date from the one most in the future to the first one.

3.3.3 Sprites and Expressions

On the final prototype we decided to make our character sprites, the images that would make the Tutors' representation for the user, from an expressions database by Cohn-Kanade[54] obtained through a student's license in order to have more natural and standardized expressions for our tutors. After we changed the tutors' avatars for the ones that were from the expressions database, aside from the default "normal" expression there were only two more available expressions for each agent: sadness and happiness. This was the trade off we compromised on for having more realistic expressions instead of a bigger variety of facial expressions that might not be recognizable by the users and leave the wrong emotional response. Each tutor had an array of Sprites for the expressions, each sprite would then be corresponded to one *enum type* object that would help us control the tutor's expressed emotion. For instance: if the tutor was supposed to change their expression to show they were feeling said we would make the Agent's expression enum variable become "SAD" which its numeric value would then correspond to the array index of the sprite correspondent to the sad expression. We found

this decision key for our application and for the exploration work that we wanted to achieve because if the agents' expressions we chose for each case were not standard expressions of an emotion and chosen solely on our interpretation of the image they could lead to wrong interpretations from the part of the users and also if the expressions were not believable enough the users might not take the tutors as serious as we want the users to take them. We found this change fundamental for our application and for the exploration work that we wanted to achieve because if the agents' expressions we chose for each case were not standard expressions of an emotion and chosen solely on our interpretation of the image they could lead to wrong interpretations from the part of the users and also if the expressions were not believable enough the users might not take the tutors as serious as we want the users to take them.

3.3.4 Integration

The last big hurdle on the implementation side was the functionality of making a study plan with the user by accessing the Google calendar present in their own smartphone. This was made using android plug-ins on the Unity3d C# scripts as alternate functionalities for the functions while they are operating on a device hosting the Android operating system. We implemented this measure in order to have the integration we mentioned in previous chapters so we could augment the positive experience felt by the users, simplify that same experience with familiar concepts and take advantage of programming for the Android operating system.

3.3.5 Design

Designwise we decided to keep with the metaphor and to express all dialogues as speech bubbles similarly to comic books. We chose this means of representation for two main reasons: on the one hand it follows our inspiration on basing our dialogue from the game *Oxenfree*, and on the other hand the speech bubble's design can be easily manipulated to convey emotions in a manner other than the tutor's facial expressions, for example a spiked balloon can convey anger or excitement depending on the context. Each tutor uses different color speech bubbles in order to identify whose line it is. The male tutor speaks with a brown speech bubble, the female tutor speaks with a yellow speech bubble and as for the user's lines they are all with green speech bubbles.

Aside from the main interaction screen we had other two menus that were used to organize information and for it to be easier for the users to reach this information without always having to talk to the agents about it. The first menu had just a list of the courses the user was taking and would then redirect the user to a menu for the selected course.

This individual course menu had the changeable metrics we have previously mentioned in the metaphor: *Confidence*, *Motivation* represented by sliders that were set by the tutors themselves but could be changed by the user if they so desire. After receiving the input of an evaluation grade, the respective course's metrics along with the actual grade on that evaluation would weigh on the agent's decision of a response. As such, on a course that is less important, harder for the student and that the student does not like even the more exigent tutor will be more permissive of a lower passing grade,

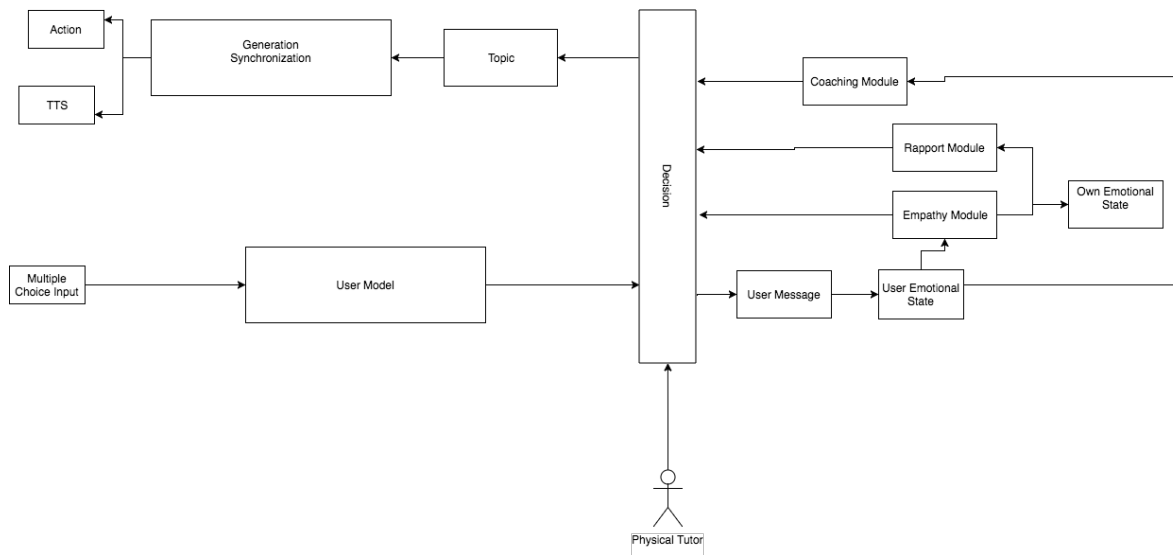


Figure 3.9: Each agent's architecture

obviously the inverse also happens. *Importance* is also represented in the sliders but it should be a locked slider in a future iteration of the prototype.

On the menu, below the metrics is the aforementioned list of *Checkpoints*.

Lastly, on the design front, after some testing we felt like we needed to provide some kind of context for the agents so that the users would intuitively realize they were using an educative app. As mentioned by one of the testers in the MoJo "These are just floating heads, at a glance I do not understand why they are here or what they want to do for me". As such, we decided to make the background of the main interaction screen as some sort of an office scenario.

On what the characters are concerned we decided to keep the initial idea in which one of them, the male, is more exigent, more pessimistic and more sarcastic and opposed to that the female is kinder, gentler, softer, more permissive and more straightforward. This clash of personality among both tutors can provide several opportunities to break the awkwardness of the interaction.

3.3.6 Sound

The TTS module was done with pre-recorded sentences using an offline version of the Ivona TTS software we acquired so that the agents could say what they wanted to say in emotionless words instead of emotional gibberish devoid of any content. We believe the TTS to be important while interacting in a quiet environment because it provides hearing help to the user. As such, the user does not need to be as focused on reading the agents' lines and can pay attention to other details in the interaction such as the agents' expressions.

3.3.7 Architecture

Figure 3.9 describes each agent's interaction loop with the user. This loop starts with an input given by the user as a choice of one of the possible choices colored green. This action gets saved as a user message and will be pondered upon in order to take a decision. This message will, if needed,

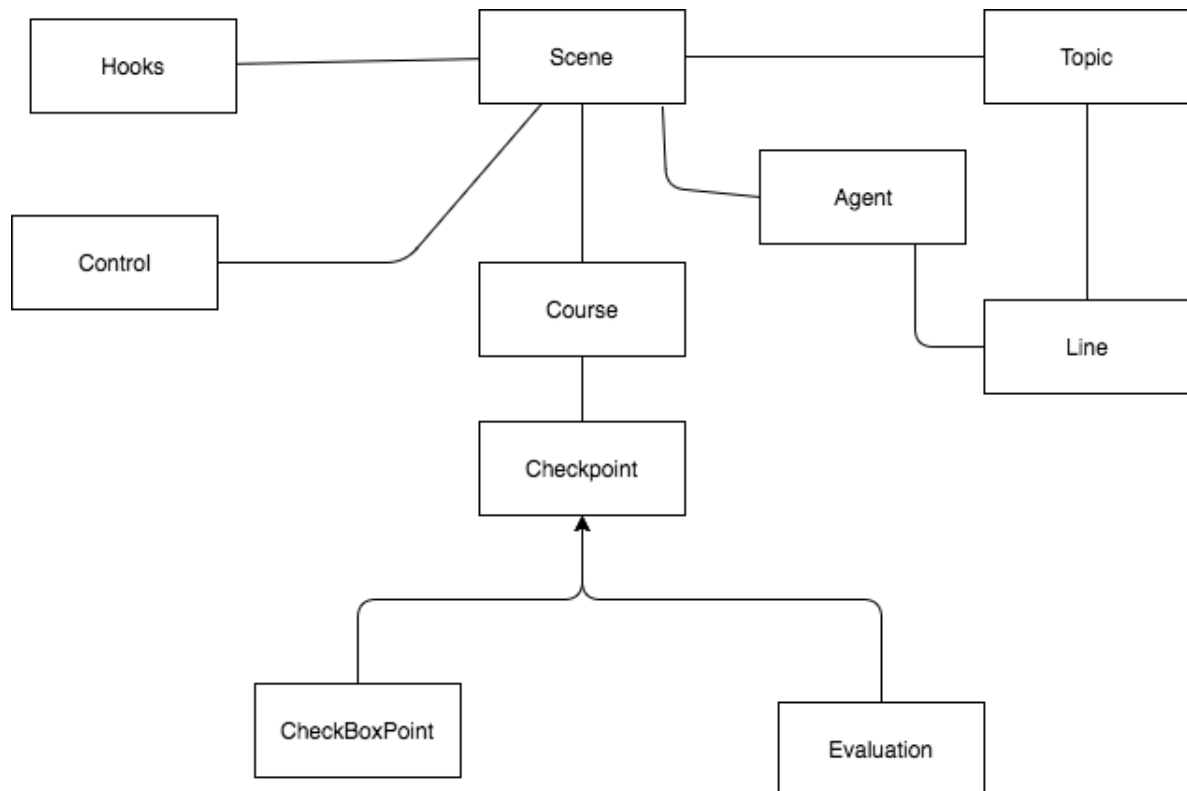


Figure 3.10: Virtual Tutor's class diagram

alter the user's perceived emotional state. That user's perceived emotional state will be transported to the *empathy module* that will combine the user's emotional state with the agent's own personality and if needed it will cause an alteration on the agent's emotional state. The user's emotional state will also be taken into account while doing coaching activities, in this case defining the optimal required time for the user to study for that next week, as we assume that the student's results are a manifestation of their own emotional state. As it is the *Coaching Module* will also contribute to the final decision based on the user's emotional state. The user's expression of it's own emotional state is decided by the *Rapport Module* and one of the facial expressions is chosen in order to express it, as well as other factors to build rapport, on a future time nods and smiles to acknowledge the conversation that is happening. As such, the user's message passes through the three modules and these three modules each take a weight on the decision of a topic of answer. A topic, as previously explained, is a set of lines, that each have their content and the agent that speaks them, the possible inputs for the user's response and all of the sentences timings. This topic will then be expressed through TTS and if the expression chosen by the *Rapport Module* is different than the one the agent is already making, the agent's expression will change with a small animation.

To note that most of this behavior is still scripted and not automated because this was the app basis to support our research.

3.3.8 Class Diagram

The *Scene* class is a manager of sorts that receives two instances of *Agent*, one for each of the Virtual Tutors; An array of *Course* instances that will determine the courses the student is attending; a speech *Topic*; *Controls* and *Hooks* for the calendar, agents's expressions and bubbles and for the splash screen.

Each *Agent* will then be associated to a sprite and according to its emotional state the sprite will change in order to show a different emotion. These different possible emotional states are all stroed in an array of enumerates.

The *Course* object will store various informations that make sense in the context of the course, such as its name, the informations related to the sliders of confidence, motivation and importance and a Dictionary of the *Checkpoints* on that course.

These *Checkpoints* can be of two kinds: *CheckBoxPoint* or *Evaluations*. The *CheckBoxPoint* relate to small objectives to be achieved throughout the course that are set by the responsible professor beforehand and can only have a boolean value on their checked state, they are represented as check boxes. *Evaluations* are associated to a *score* level instead of a boolean and relate to graded tests or projects.

The *Topic* remotes to a set of the interaction, a conversational set until the user inputs a message, it is a set of *Lines* and the possible input choices that are available for the user for that part.

A *Line* is a set of sentences, the *Agent* who speaks them and their timings of appearances.

3.3.9 Concluding remarks

In order to reach our final prototype we had to go through various phases and work on several fronts. We started with an idea of having two Virtual Agents with complementary personalities cooperating to provide support to a student in need through coaching. For coaching purposes we decided that the student's relationship with each course would be specified by two variables, the student's confidence in their prowess in the course and the student's motivation for learning the course's contents. We decided to introduce the variable related to the course's importance in the general context of the degree to provide a higher management of the tutors' strictness in the different courses.

After two moments of usability tests to polish this prototype, we reached our final version of the prototype with:

- Adequate dialogue for the intended target.
- Sprites obtained through an expression database, that although they are less expressive, their expressions are validated to an emotion.
- Well-timed timings for the appearance of dialogue and change of emotional expressions, integration with *Google Calendar* to schedule study plans and *Ivona* for Text-to-speech purposes.
- A defined color scheme for each agent's speech bubble and for the user's input choices.
- Scenarios for reacting when the student is not paying attention to the tutors.

- A course menu with checkpoints that may be associated with a numeric value representing grade or may be associated to a boolean value representing completion.

4

The Experiment

Contents

4.1 Experimental Protocol	I
-------------------------------------	---

This chapter will show how we decided to divide the three main components we found for a Virtual Tutor: empathy, rapport and the relationship, each in two different behaviors for a Virtual Tutor and how we converted those behaviors into film in order to be tested by the users. You will also find the experimental protocol and setup as well as the results of said experiment.

4.1 Experimental Protocol

At this moment we had the app ready to start our exploratory work on the three domains that we had previously investigated: the five phases of a relationship proposed by [18], building rapport and generating empathy. After careful consideration we decided to divide each of these three main aspects of our research into two behavioral components each: for the relationship theme we chose the acquaintance and maintenance phases, for the rapport theme we chose the positivity enforcing strategy of self-disclosure and the coordination and mutual attention enforcing strategy of back-channeling and for the empathy theme the posture mimicry through emotional mimicry and perspective taking. These 6 components were then presented to the users as a short video and the users were able to grade them according to the importance they thought each component had, and also they were required to provide small description of maximum three adjectives to each observed behavior.

The idea was to record six clips, one for each of the behavioral components we had chosen. For that we had to make some alterations to our agents' functionalities in order to isolate each of the behavioral components we wanted to explore and we just emphasized that specific behavioral component. We made these alterations mostly on the dialogue script and we normalized the scenario. The scenarios would all have the same common points: the user would login and interact with the tutors, after pleasantries and greetings, the tutors would warn the user that their Algebra's test grades had already been published, the user would then see it and notice that they had gotten a 7.0 out of 20.0 for the most recent Algebra test, the video will then show both agents' reactions to the result depending on the behavior that we are trying to emphasize, different behaviors will have of course different reactions although the context is always the same. The clips were recorded on an Android screen capture with emphasis on the user's cursor so that the participants in the experience could see clearly the choices made.

4.1.1 Relationship stages - Acquaintance and Maintenance

For the relationships we focused on the phases of acquaintance and maintenance because we wanted to see how the users would react as to being treated more or less familiarly by a tutor that would impose this familiarity in their manners of speech, as such we wanted to know which kind of familiarity degree would ensure a more pleasant feeling or reaction from the user's part or if there was any difference in the way the relationship was emulated. As such we chose these two degrees: acquaintances and good friends.

On the acquaintance clip there needed to be a higher degree of positivity or pleasantness from the part of the agents' dialogue than if the agents and the user were already friends. The language

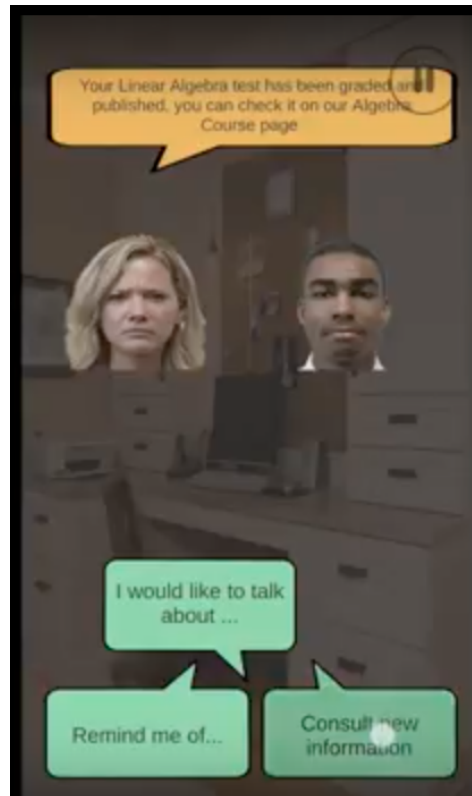


Figure 4.1: Example of one of the clips interactions with a cursor.

was also more formal and the agents were also careful to be respectful and mindful of anything they said as not to offend or insinuate a closer relationship without any previous trust.

On the other hand on the maintenance phase clip the user and the Virtual Tutors had a relationship on the maintenance stage, that is to say they were friends and the relationship needed to be maintained. Nothing was said directly to the user about that relationship state and they would have to infer it on their own while watching the clip. This relationship level transposed a little bit of extra dialogue on which the agents recalled a past interaction with the user, as this is one of the maintenance strategies proposed by Levinger[18], as well as a less formal language and less efforts on trying to be pleasant to the user because they already are friends, that means the agents did not try their hardest not to be misinterpreted on the grounds that there was already a close and functioning relationship. On this approach, the agents will had a few less serious remarks due to their closeness to the user.

4.1.2 Rapport

The two clips that followed we chose the two most important rapport building factors in our context and we hope to know which of these is more important for the context of Virtual Tutoring: mutual attention through back-channeling or positivity from self-disclosure. One of the clips was focused on positivity through the strategy of self-disclosure and the other mixed coordination and mutual attention by a back-channeling through dialogue. Although the research done in the state of the art section shows us that positivity, mutual attention and coordination are fundamental in order to build rapport and have a harmonious interaction flow we want to discover which of these two factors is

more important in this determined context, that means which one people notice more on the context of education.

On the first clip of this theme we tried to build rapport by using the strategy of self-disclosing the agents' emotions and information. In this clip there was an emphasis on self-disclosure of the agents' own thoughts, emotions and experiences on what the user has inputted and also their emotional reactions to the user's grades. Research in the state of the art on rapport tells us that self-disclosure is one of the best strategies to increase positivity in the interaction[27]. With it we wanted the agents to be regarded as friendlier and more pleasant to interact with. Our objective is for users to have a good time while interacting with the agents and consequently keep the number of interactions rising.

On the other hand, the second video on this theme only emphasized mutual attention. In our specific case mutual attention was achieved through back-channeling in the Virtual Tutor as a means that manages responses to the inputs given by the user or the other agent's sentences. The back-channeling was only of the verbal kind, using connective expressions like "really?", "hmmm" and "what do you think?". In this interaction the agents spoke more amongst themselves than with the user. With this approach we wanted to give the feeling that both agents were paying attention to each other and the user while they were interacting.

4.1.3 Empathy - Posture mimicry and Perspective taking

The last theme we wanted to test was empathy. Based on the state of the art we gathered that empathy is achieved especially on two levels[19][44]: posture mimicry and perspective taking. Each of the clips focused on one specific aspect of empathy generation and will show these factors' importance on generating empathy with the user.

On the first clip of this theme we tried to generate empathy by using the strategy of posture mimicry. As such, in this clip the agents mimicked the user's emotion based on what had been gathered from the user's affective model. The agents showed emotions that were not their own but convey them differently in their speech. Note that this emotion does not take into account the user's own predictions and is just pure emotion without any advice or ways to cheer the user up, so they do not take the user's full perspective, they just mimicked the perceived emotion.

As opposed to posture mimicry the second clip brought the user's emotions and emotional state more into the interaction loop as the agents' own feelings were influenced by the feelings they infer the user is feeling throughout the interaction. As such, it is said that the agents are taking the user's perspective and seeing everything that happens to the user as if it were happening to themselves. The agents felt like they failed when they saw the 7 on the score sheet, they also reacted differently according to their own personalities and proposed ways to make the situation better and to improve the results on the next evaluation moment. We hope this will create an empathic bond between the agents and the user.

4.1.4 Procedures

All of these videos were recorded and submitted to six different Google Form questionnaire that were only different in the order the videos were shown as to not make unconscious prejudicial thoughts while watching each video making sure that after a big number of participants the results we would gather were not dependent on the order the videos were shown and the answers we would gather would be balanced. We used a web page that would redirect each user to one of the respective questionnaires with the use of Roman Square strategy to avoid any kind of bias, for example user 1 would go to questionnaire 1 and user 2 to questionnaire 2, user 7 would go to questionnaire 1 again and so on so forth.

Each questionnaire was divided in two parts: demographic questions and the video analysis. On the demographic questions we asked the standard questions of age, gender, the participant's familiarity with apps, the participant's familiarity with virtual assistants, the participant's familiarity with the calendar app on their smartphone and the participant's familiarity with NPC's (Non-playable characters in games). On the video analysis part the participant would watch each video and after each the participant would answer two questions: on a scale from 1 to 9 what was the importance of the observed behavior and with three adjectives we asked them to describe the agents' behavior.

With these questions we wanted to discover which of these behaviors were deemed more important for a Virtual Tutor, why those behaviors were that important and what the participants felt when dealing with these behaviors from the part of the agents.

The experience took place online from July 31 to August 20. In the subsections below we will talk about the participants' demographic results followed by the results of the experiment.

5

Results from the Experiment

Contents

5.1	Participants	lvi
5.2	Results of the experiment	lvii
5.3	Statistic Tests and respective results	lix

5.1 Participants

5.1.1 Age

In relation to the participants' age we gathered the following results: As you can see in figure B.1, our participants' age varied from 15 to 42 years of age. With an average value of 23.60 and a mode of 24 years of age. This means the most of our participants were around the intended target audience for the *Virtual Tutor* app. The majority of the participants were from 21 to 25 years of age with some outliers, being the more different outliers a 42 year-old and a 15 year-old participants.

5.1.1.A Gender

In relation to the participants' gender we gathered the following results:

According to the pie chart in figure B.3 in our experiment we had 18 female participants and 12 male participants which accounts to 60% female participation and a 40% male participation.

5.1.1.B Student

In relation to whether the participants were students in university, students elsewhere or not students, we gathered the following results:

In this experiment 18 of the participants were University students and 5 were students elsewhere. That means that roughly 83% of the participants were students, and probably could deal with the same problems we want the Virtual Tutors to help with as you can see in the pie chart in figure B.5.

5.1.1.C Participants' other demographic data

Aside from asking each participant's gender, age and whether or not they were a student in University we asked four more questions about each participant's demographic background that we deemed relevant to their point of view regarding their experience with the Virtual Tutor's videos. These four questions were:

- How often do you use mobile apps?
- How often do you use the calendar app on your smartphone?
- Have you ever had experience with virtual assistants like Apple's "Siri", Microsoft's "Cortana" or Amazon's "Alexa"?
- How often do you play games that require you to interact with Non-Player Characters (virtual characters that you don't control)?

As stated above, these control questions were used to later check whether or not the user's experience and familiarity with each factor had any relevance into their reactions to the virtual tutors.

On the above figure B.6 is shown that 22 of the 30 people that answered the questionnaire use mobile apps every day which is around 73.3 %, with only one person of the 30 answering that they never use any kind of mobile apps.

Figure B.7 shows that the majority of the participants occasionally use their own calendar apps in their smartphones, 70% to be exact, whereas 10% never uses it and 20 % uses it every day. This means 90% of the sample is familiar with the calendar app in their smartphones which we can extrapolate that most people are. Meaning most people will be familiar with the way the Virtual Tutor app handles the management of study plans with the user.

Figure B.8 shows that none of the participants in the sample interacts with any of the Virtual Assistants given as examples every day and that only 10 of the 30 participants actually interact with them. That is 66.7 % of the people in the sample that never interact with a virtual assistant.

These questions' answers were slightly different than the previous ones. Instead of the previous answers that were based on the frequency of interaction, we chose to put these answers based on the preference of games in which the participant had to interact with NPC's, as opposed to games where the user did not have to interact with NPC's. In figure B.9 we can see that to 17 of the 30 participants, which corresponds to 56.7% it does not pose a relevant difference whether or not they have to interact with NPC's in a game. However, the percentage of people that prefer to interact with NPC's in a game was really close to the percentage of people that prefer not to interact with NPC's in a game. We speculated that the people that do not like to interact with NPC's would be more focused on the functionality of the app than with how the agents would be treating them.

5.2 Results of the experiment

As stated in the section related to the questionnaires themselves, each video was correspondent to one of the six categories we wanted to study. And to each video we recovered the characteristic's importance and 3 adjectives to the tutor's behavior when we emphasized that characteristic. This section will concern the results directly inferred from the answers to the questionnaires without recurring to any statistic test whatsoever, as it will only mention the studied variables.

5.2.1 Importance

The information regarding each video's characteristic's importance was the following:

Analyzing figure 5.1 we gathered that the behavior that shows the highest mode and median values among the values of the importance of the shown behaviors in the questionnaires was the empathy strategy of perspective taking. It had a mode value of 8 out of 10 and a median value of 7 out of 10. Although the behaviors relative to the acquaintance strategy of the relationship and the rapport strategy of back channeling had the same mode, they had slightly lower medians, meaning that these behaviors were, in frequency, voted as slightly less important by the public as seen on the respective appendix C. The behavior with the weakest mode was the one regarding the empathy strategy of emotional mimicry and the one with the lowest median was the behavior regarding the rapport strategy of self-disclosure. The full frequency tables are available as appendixes in the respective chapter C.

		Statistics					
		Acquaintance_Importance	Maintenance_Importance	SelfDisc_Importance	BackChannel_Importance	Mimicry_Importance	Perspective_Importance
N	Valid	30	30	30	30	30	30
	Missing	1	1	1	1	1	1
Median		6,50	7,00	5,00	6,50	6,00	7,00
Mode		7 ^a	7	5 ^a	7 ^a	3 ^a	8

a. Multiple modes exist. The smallest value is shown

Frequency Table

		Acquaintance_Importance			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not important at all	1	3,2	3,3	3,3
	2	2	6,5	6,7	10,0
	3	1	3,2	3,3	13,3
	4	1	3,2	3,3	16,7
	5	5	16,1	16,7	33,3
	6	5	16,1	16,7	50,0
	7	6	19,4	20,0	70,0
	8	6	19,4	20,0	90,0
	Fundamental	3	9,7	10,0	100,0
	Total	30	96,8	100,0	
Missing	System	1	3,2		
Total		31	100,0		

Figure 5.1: Statistic information on the Importance variable in the six videos

5.2.2 Adjectives

Regarding the three adjectives to each video, since many words have the same meaning and some do not have any relevance towards our conclusion we decided to separate these adjectives in two ways. First we separated the positive qualifications and the negative qualifications for each tutor's behavior in each video. Then we gathered all of the adjectives again and separated them into different categories: adjectives that described the tutor as showing some kind of emotion; adjectives that describe the tutor as being serious, formal, professional or anything similar; adjectives that described the tutor as being friendly, dynamic, modern or anything similar; adjectives that described the tutor as helpful, supportive or anything similar.

As we can see from Figure D.1, the most positively qualified tutor per participant on average was the one that emphasized perspective taking and the one with the least positive adjectives per participant on average was the one that emphasized the self disclosure strategy.

As we can see from D.2, the most negatively qualified tutor per participant on average was the one that emphasized the self disclosure strategy and the one with the least negative adjectives per participant on average was the one that emphasized perspective taking.

As we can see from Figure D.3, the tutor that was most commented on as a sentient tutor per participant on average was the one that emphasized the perspective taking strategy and the one with the least qualifications of being a sentient tutor per participant on average was the one that emphasized the relationship strategy of treating the user like an acquaintance.

As we can see from Figure D.4, the tutor that was most commented on as a serious tutor per participant on average was the one that emphasized the relationship strategy of treating the user like an acquaintance and the one with the least qualifications of being a serious tutor per participant on average was the one that emphasized the rapport strategy of self-disclosure

As we can see from Figure D.5, the tutor that was most commented on as a friendly tutor per

participant on average was the one that emphasized the relationship strategy of maintaining that relationship and the ones with the least qualifications of being friendly tutors per participant on average were the ones that emphasized empathy strategies of mimicry and perspective taking.

As we can see from Figure D.6, the tutors that were most commented on as supportive or helpful tutors per participant on average were the one that emphasized the maintenance of the relationship and the one that emphasized the empathy strategy of perspective taking, the one with the least qualifications of being a supportive or a helpful tutor per participant on average was the one that emphasized the rapport strategy of self-disclosure.

Having gathered all of the results from the online questionnaires we decided to analyze them in order to discover for sure whether any of the six factors that we isolated: the acquaintance phase of the relationship, the maintenance phase of the relationship, the self disclosure strategy for rapport to increase positivity, the back channeling behavior for rapport in order to increase coordination and mutual attention, the posture mimicry and the perspective strategies for generating empathy were in fact statistically relevant to ascertain if there was a relative importance among these factors so we could define which was more important in the educational context of a Virtual Tutor.

5.3 Statistic Tests and respective results

This section will present all the statistic tests and conclusions taken from them according to the results taken from the variables in the previous section.

5.3.1 Shapiro-Wilk test to check normality

We started by doing a Shapiro-Wilk test in order to realize whether or not the data of each video's importance had a normal distribution. Since the data did not have a normal distribution in at least one of those cases we would be forced to make non-parametric tests to ascertain whether or not there were relevant variants in the data for each of these cases. In the figures below we can see the results for each of the performed Shapiro-Wilk tests:

As we can see from figure E.1, the value of p on the Shapiro-Wilk's test was lower than 0.05, indicating that the values for the video correspondent to the acquaintance phase of the relationship do not have a normal distribution.

As we can see from figure E.2, the value of p on the Shapiro-Wilk's test was lower than 0.01, indicating that the values for the video correspondent to the maintenance phase of the relationship do not have a normal distribution.

As we can see from figure E.3, the value of p on the Shapiro-Wilk's test was 0.246, indicating that the values for the video correspondent to the rapport strategy of self disclosure for the increase of positivity have a normal distribution.

As we can see from figure E.4, the value of p on the Shapiro-Wilk's test was lower than 0.05, indicating that the values for the video correspondent to the rapport strategy of back channeling for the increase of coordination and mutual attention do not have a normal distribution.

As we can see from figure E.5, the value of p on the Shapiro-Wilk's test was lower than 0.05, indicating that the values for the video correspondent to the empathy generation strategy of posture mimicry through emotional mimicking do not have a normal distribution.

As we can see from figure E.6, the value of p on the Shapiro-Wilk's test was lower than 0.01, indicating that the values for the video correspondent to the empathy generation strategy of perspective taking do not have a normal distribution.

5.3.2 Friedman General test

As such, the only data that had a normal distribution was the result from the video that emphasized the self disclosure strategy of building rapport by increasing positivity, we had to do non-parametric tests in order to find out whether any of the behaviors shown in the videos had a statistic relevant greater importance than any of the others. In order to know for sure, we performed a Friedman test with all of the answers referent to all of the six behaviors' importance for the participants.

As we can see from figure E.7, the value of p on the Friedman test was higher than 0.05, meaning that the values for the behavior's importances shown in the videos were considered equally important.

5.3.3 Wilcoxon signed-rank test in pairs

Even if in general all of the behaviors' importances were considered equally important, we decided to compare each video to check between every two behaviors if one was considered more important than the other. For that we did a series of Wilcoxon signed-rank tests with every possible combination of two videos. As we can see from figure F.1 in the appendix F, all the value of p on the Wilcoxon signed rank tests were higher than 0.05, meaning there are no relevant statistical differences between the importance of any pair of behaviors in the whole behavior set. Meaning there were no behavior was statistically more relevant than any other behavior when comparing the two individually.

5.3.4 Friedman test with a split the dataset

With all of the Wilcoxon signed-rank tests performed we could finally conclude that generally speaking with all our samples there was no statistical difference between any importance of any of the shown behaviors. Again since this took into account all the data, we decided we could deepen the search for differences in the shown behavior's importances but now splitting the sample by the demographic groups trying to find patterns.

The first demographic factor we focused on was gender, however the Friedman tests indicated no difference between the importances chosen by these groups. G.1G.2

After splitting the dataset by gender, we decided to focus on the age difference. We separated the dataset on people that were older than 24 years old, people that were younger than 24 years old and people that were 24 years old. We chose the age of 24 as a separator because it was the mode for the values of age and had a significant amount of people in that category, however the Friedman tests indicated no difference in the importances chosen by these groups. G.3G.4G.5

After splitting the dataset by age, we decided to focus on the fact of whether or not the participant was a student. For this case we decided to split the dataset in University students and not University students. We chose this fact as a divider because University students are the main users of the Virtual Tutor app and they are more familiarized with the worries and challenges of an academic life in University, however the Friedman tests indicated no difference between the importances chosen by these groups. G.14G.15

After splitting the dataset by the fact of whether or not the participant was a University student, we decided to focus on the frequency of the participants use of their smartphone apps. Since our data had only one participant that answered they never used smartphone apps, we divided this dataset between people that used smartphone apps everyday and people that did not use smartphone apps every day, however the Friedman tests indicated no difference between the importances chosen by these groups.G.12 G.13

After splitting the dataset by the frequency of usage of smartphone apps, we decided to focus on the frequency of the participants use of the calendar app in their smartphone. The calendar is an important feature for the Virtual Tutor as we planned the scheduling of the study plans with the Google Calendar app. However the familiarity with the calendar is not that much relevant for the participants experience of the videos because we did not show them a scheduling of a study plan. For this factor, we split the dataset in three parts with the frequencies of "every day", "occasionally", and "never", which were the same for the answers to the demographic question on the form,however the Friedman tests indicated no difference between the importances chosen by these groups, however the Friedman tests indicated no difference between the importances chosen by these groups. G.7G.6G.8

After splitting the dataset by the frequency of usage of their calendar apps, we decided to focus on the frequency of the participants interaction with Virtual Assistants like Apple's Siri, Microsoft's Cortana or Amazon's Alexa. Since we are testing the importance of behaviors in Virtual Tutors which are some sort of Virtual Assistant for education purposes we wanted to know whether the familiarity with these kinds of agent had an impact on choosing a more important behavior. In this demographic, since we did not have any participant that answered that they interacted with Virtual Assistants every day we decided to do two tests: one with people that had never interacted with a Virtual Assistant and another with people that occasionally interacted with a Virtual Assistant, however the Friedman tests indicated no difference between the importances chosen by these groups.G.17G.16

The last split of dataset we wanted to make was on the demographic question of whether the participant preferred a game in which they had to interact with NPCs, non-player characters or characters that are not being controlled by players. We thought this an important value that could have their own importance because the Virtual Tutors are some sort of NPC with which the user interacts. Perhaps the familiarity with these kind of characters could lead to a hierarchy in the importance of their characteristics, however the Friedman tests indicated no difference between the importances chosen by these groups. G.9 G.10 G.11

5.3.5 Discussion of Results

After doing the Friedman tests with all the possible demographic groups data we came to the conclusion that there was no behavior that was statistically more important than any other one. So we can assume that either all of the behaviors in the videos were equally important for the making of a Virtual Tutor, or that the videos did not pass the intended idea due to not accurate expression, showing more than one behavior because it is hard to isolate one type of behavior separately. By getting the statistics of the adjectives for each video we showed in the results and comparing them we see that the behavior that emphasized the acquaintance level of the relationship was deemed as more serious than the others and its own most used adjective was related to its own seriousness as expected.

The behavior that was deemed as the most friendly one was the one that emphasized the maintenance phase of the relationship and it was also its most mentioned adjective. The one that was most mentioned as emotional was the one that used the perspective taking strategy to generate empathy, however it came really close to the other empathy generation strategy video and the self disclosure video who were the other strategies that had to deal with emotions.

However, we can not say that this data alone can clear the insecurity of the videos being explanatory by themselves for several factors: all of the tutors behaviors received a considerable amount of negative adjectives that some happened because the participants expected more from a tutor than just what it did, or they thought they were being dishonest or intrusive, in our opinion one of the factors is the ever present difference in how two people experience the same behavior that will lead to different and sometimes polar opposite reactions. Some behaviors that were not supposed to show emotion were deemed as emotional or sentient because the people felt the tutors were caring for them as they were just doing their job of tutoring, so the participants added some sort of caring persona to a unit who is responsible for another. Finally, some participants in our study did not see much difference between the behaviors shown in each of the video as some participants decided to leave similar adjectives or synonyms as answers to each video they were shown. This fact is also proven by some acquaintances directly pointing out to us after filling the questionnaires phrases like "I don't know if I did it right. They all seemed the same to me".

As a result we feel that some videos were either too short to make the participants notice the differences in them. The videos were from 40 seconds to one minute long and perhaps that duration was too short to demonstrate all of the intricacies of the selected behaviors. However, we believe longer videos may have worked against us as they might have bored the users taking into account they were six different videos. Another of our hypothesis is that the subject of the questionnaire was too vague, when we ask what the person thought of a behavior used by the agent and how important they think that behavior is without even revealing what kind of behavior we are studying it can sometimes lead to people seeing things that are not there, and we are afraid this factor may have also taken an influence with our sample. Perhaps there were too many characteristics to be tested and we decided to divide into six behaviors, perhaps it would have been better to study the difference between the importances of empathy, rapport and relationship building as just three separate behaviors in a Virtual

Tutor.

6

Conclusions and Future Work

In this document we covered our gathering of requirements and the further implementation of a prototype of a virtual tutor with two complementary personalities that use strategies of rapport, empathy and theory of relationships to better interact with a human student in order to perform tutoring and coaching, we presented early stages of the prototype to get concept validation and received positive results. After having a full interaction cycle developed, we had some usability tests done in order to discover interface and experience issues as well as concept validation with students, who are a different type of customer than the integrands of the committee to whom we presented the concept. The final version of the prototype that is described in full in this document was used as the basis of a study to identify which among rapport, empathy and the relationship, the three factors that stood as pillars of our study, was the most important in the educational context of our app.

We come back to our hypothesis defined in the beginning of this document of examining whether or not one of the six characteristics, the acquaintance and maintenance of the relationship, the rapport strategies of self-disclosure of emotions and back-channeling and the empathy generation strategies of emotional mimicry and perspective taking, based on between the user and the agent was statistically more important than the others, or if there was any specific order of importance among them. With the Friedman Tests and the Wilcoxon signed rank tests we reached the conclusion that there is no statistical difference among any of the six behaviors presented in the respective videos from the questionnaires filled by the users. However there may have been many inconsistencies in the videos that impeded us on getting accurate results for this experience.

As future work we would suggest making an in depth study on the preference of rapport as a whole, empathy as a whole and the relationship as a whole to find out if in general one of these is more important than the others. These factors should be on a more closed environment than "please explain what you thought of the tutor's behavior" because the question may be misleading. As far as the prototype goes, we would be happy if it was continued as a full app to support students throughout their university years and perhaps even younger students with some slight changes to it.

Bibliography

- [1] A. Chowanda, P. Blanchard, M. Flintham, and M. Valstar, "Erisa: Building emotionally realistic social game-agent companions," *Intelligent Virtual Agents*, 2014.
- [2] B. Henriques, "Establishing harmonious relationships between robots and humans," Ph.D. dissertation, Instituto Superior Técnico, 2016.
- [3] F. F. para a Ciência e Tecnologia, "Projectos de investigação."
- [4] H. Yan and T. Selker, "Context-aware office assistant," *Proceedings of the 5th International Conference on Intelligent User Interfaces*, 2000.
- [5] C. Dryer, "Wizards, guides and beyond: Rational and empirical methods for selecting optimal and intelligent user interface agents," *Proceedings of the 1997 Int'l Conf. on Intelligent User Interfaces*, 1997.
- [6] M. El-Najdawi and A. Stylianou, "Expert support systems: integrating ai technologies," *Communications of ACM*, vol. 36, no. 12, 1993.
- [7] F. Zambonelli, N. Jennings, and M. Wooldridge, "Developing multiagent systems: The gaia methodology," *ACM Transactions on Software Engineering and Methodology*, vol. 12, no. 3, 2003.
- [8] J. Wainer and D. Braga, "Symgroup: applying social agents in a group interaction system," *Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work*, 2001.
- [9] B. Marin, A. Hunger, S. Werner, S. Meila, and C. Schuetz, "Roles of an intelligent tutor agent in a virtual society," *Proceedings of the 2005 Symposium on Applications and the Internet*, 2005.
- [10] L. Aroyo and P. Kommers, "Preface - intelligent agents for educational computer-aided systems," *Journal of Interactive Learning Research*, vol. 10, 1999.
- [11] J. Kay, "Learner control. user modelling and user adapted interaction," *Kluwer Academic Publishers*, 2001.
- [12] S. Syed-Mustapha, "Agent mediated for intelligent conversational channel for social knowledge building in educational environment," *Proceedings of the fifth international conference on Information Technology*, pp. 533–538, 2004.

- [13] P. Chakravarthy, "The difference between coaching and mentoring," *Forbes*, 2011.
- [14] J. Renton, *Coaching and Mentoring: What they are and How to Make the Most of Them*, 1st ed., B. Press, Ed. The Economist, September 2009.
- [15] T. Bickmore and R. Picard, "Establishing and maintaining long-term human-computer relationships," *ACM Transactions on Computer-Human Interaction*, vol. 12, no. 2, 2005.
- [16] E. Berscheid and H. Reis, "Attraction and close relationships," *The Handbook of social psychology*, pp. 193–281, 1992.
- [17] E. Berscheid, "Interpersonal relationships," *Annual Review of Psychology*, vol. 45, no. 1, pp. 79–129, 1994.
- [18] G. Levinger, "Development and change." *Close Relationships*, pp. 315–359, 1983.
- [19] I. Leite, "Long-term interactions with empathic social robots," Ph.D. dissertation, Instituto Superior Técnico, 2013.
- [20] L. Stafford, M. Dainton, and S. Haas, "Measuring routine and strategic relational maintenance: Scale revision, sex versus gender roles and the prediction of relational characteristics," *Communication Monographs*, vol. 67, no. 3, pp. 306–323, 2000.
- [21] T. Bickmore, "Relational agents: Effecting change through human-computer relationships," Ph.D. dissertation, Massachusetts Institute of Technology, 2003.
- [22] M. Ochs, N. Sabouret, and V. Corruble, "Simulation of the dynamics of nonplayer characters' emotions and social relations in games," *IEEE Transactions on Computational Intelligence and AI in Games*, pp. 356–361, 2013.
- [23] T. Almaev and M. Valstar, "Local gabor binary patterns from three orthogonal planes for automatic facial expression recognition," *Affective COmputing and Intelligent Interaction*, 2013.
- [24] S. Kopp, B. Krenn, S. Marsella, and et al., "Towards a common framework for multimodal generation: The behavior markup language," *Int'l Conf. Intelligent Virtual Agents*, pp. 21–23, 2006.
- [25] G. Saucier and L. Goldberg, "The language of personality: Lexical perspectives on the ve-factor model," *The ve-factor Model of personality: Theoretical perspectives*, pp. 21–50, 1996.
- [26] J. Gracht, N. Wang, J. Greten, E. Fast, and R. Duffy, "Creating rapport with virtual agents," *Intelligent Virtual Agents*, 2007.
- [27] H. Spencer-Oatey, "(im)politeness, face and perceptions of rapport: Unpacking their bases and interrelationships," *Politeness Research*, 2005.
- [28] Tickle-Dengen and R. Rosenthal, "The nature of rapport and its nonverbal correlates," *Psychological Inquiry*, vol. 1, no. 4, pp. 285–293, 1990.

- [29] A. Drolet and M. Morris, "Rapport in conflict resolution: accounting for how face-to-face contact fosters mutual cooperation in mixed-motive conflicts," *Experimental Social Psychology*, vol. 36, pp. 26–50, 2000.
- [30] J. Cogger, "Are you a skilled interviewer?" *Personnel Journal*, 1982.
- [31] P. Tsui and G. Schultz, "Failure of rapport: Why psychotherapeutic engagement fails in the treatment of asian clients," *American Journal of Orthopsychiatry*, pp. 90–104, 1985.
- [32] D. Fuchs, "Examiner familiarity effects on test performance: implications for training and practice," *Early Childhood Education*, 1987.
- [33] M. Burns, "Rapport and relationships: The basis of child care," *Journal of Child Care*, vol. 2, pp. 47–57, 1984.
- [34] J. G. et al., "Virtual rapport," *6th international Conference on Intelligent Virtual Agents*, 2006.
- [35] —, "Can virtual humans be more engaging than real ones?" *12th International Conference on Human-Computer Interaction*, 2007.
- [36] J. Smith, "Grandchair: Conversational collection of family sotries," *MediaLab*, 2000.
- [37] J. Bailenson and N. Yee, "Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments," *Psychological Science*, pp. 814–819, 2005.
- [38] W. Burlison and R. Picard, "Evidence for gender specific approaches to the development of emotionally intelligent learning companions," *IEEE Intelligent Systems, Special issue on Educational Systems*, 2007.
- [39] R. Zhao, A. Papangelis, and J. Cassell, "Towards a dyadic computational model of rapport management for human-virtual agent interaction," *Intelligent Virtual Agents*, pp. 514–527, 2014.
- [40] B. Mutlu, J. Forlizzi, and J. Hodgins, "A storytelling robot: Modeling and evaluation of human-like gaze behavior," *Proceedings of the 2006 6th IEEE-RAS International Conference on Humanoid Robots*, 2006.
- [41] S. Andrist, B. Mutlu, and A. Tapus, "Look like me: Matching robot personality via gaze to increase motivation," *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015.
- [42] R. Niewiadomski, E. Bevacqua, and C. Pelachaud, "Greta: An interactive expressive eca system," *Aamas*, 2009.
- [43] S. Kang and J. Gratch, "People like virtual counselors that highly-disclose about themselves," *Studies in health technology and informatics*, 2011.
- [44] C. Anderson and D. Keltner, "The role of empathy in the formation and maintenance of social bonds." *Behavioural and Brain Sciences*, vol. 25, no. 1, pp. 21–22, 2002.

- [45] F. D. Vignemont and T. Singer, "The empathic brain: how, when and why?" *Trends in Cognitive Sciences*, vol. 10, no. 10, pp. 435–441, 2006.
- [46] M. Hoffman, "Empathy and moral development: Implications for caring and justice," *Cambridge Univ Press*, 2001.
- [47] S. Preston and F. de Waal, "Empathy: Its ultimate and proximate bases." *Behavioural and Brain Sciences*, vol. 25, no. 1, pp. 1–20, 2002.
- [48] S. Brave, C. Nass, and K. Hiutchinson, "Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent," *International Journal of Human-Computer Studies*, 2005.
- [49] A. Paiva, J. Dias, D. Sobral, R. Aylett, P. Sobreperez, S. Woods, C. Zoll, and L. Hall, "Caring for agents and agents that care: Building empathic relations with synthetic agents," *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems*, vol. 1, pp. 194–201, 2004.
- [50] A. Paiva, J. Dias, D. Sobral, R. Aylett, S. Woods, L. Hall, and C. Zoll, "Learning by feeling: Evoking empathy with synthetic characters," *Applied Artificial Intelligence*, 2005.
- [51] N. Lessmann, S. Kopp, and I. Wachsmuth, "Situating interaction with a virtual human - perception, action and cognition," *Situated Communication*, 2006.
- [52] H. Boukricha, I. Wachsmuth, A. Hofstätter, and K. Grammer, "Pleasure arousal - dominance driven facial expression simulation," *3rd International conference on Affective Computing and Intelligent Interaction*, 2009.
- [53] J. Z. Jake Knapp, *Sprint: Solve big problems and test new ideas in just 5 days*, 1st ed., Simon and Schuster, Eds. Simon and schuster, 2016.
- [54] (2017, October). [Online]. Available: <http://www.pitt.edu/~emotion/ck-spread.htm>



The Questionnaire

Questionnaire VT

The following is a questionnaire to evaluate the needed characteristics for a pair of Virtual Tutors that act as your coaches. A Virtual Tutor is a virtual agent that will help you through a school or university semester. You will watch 6 videos that each represent one behavior characteristic we believe to be fundamental for a virtual agent on this context. Help us understand which of these you value more in a Virtual Tutor by answering the questions honestly

***Obrigatório**

1. 1 - Age: *

2. 2- Gender *

Marcar apenas uma oval.

Female

Male

3. 3- Are you a student? *

Marcar apenas uma oval.

Yes, on University

No

Outra: _____

4. 4- How often do you use mobile apps? *

Marcar apenas uma oval.

I never use mobile apps

I occasionally use mobile apps

I use mobile apps every day

5. 5- How often do you use the calendar app on your smartphone? *

Marcar apenas uma oval.

I never use my calendar app

I occasionally use my calendar app

I use the calendar app every day

6. 6- Have you ever had experience with virtual assistants like Apple's "Siri", Microsoft's "Cortana" or Amazon's "Alexa"? *

Marcar apenas uma oval.

I don't normally use apps where I have to interact with virtual assistants

I sometimes interact with virtual assistants

I interact with virtual assistants every day

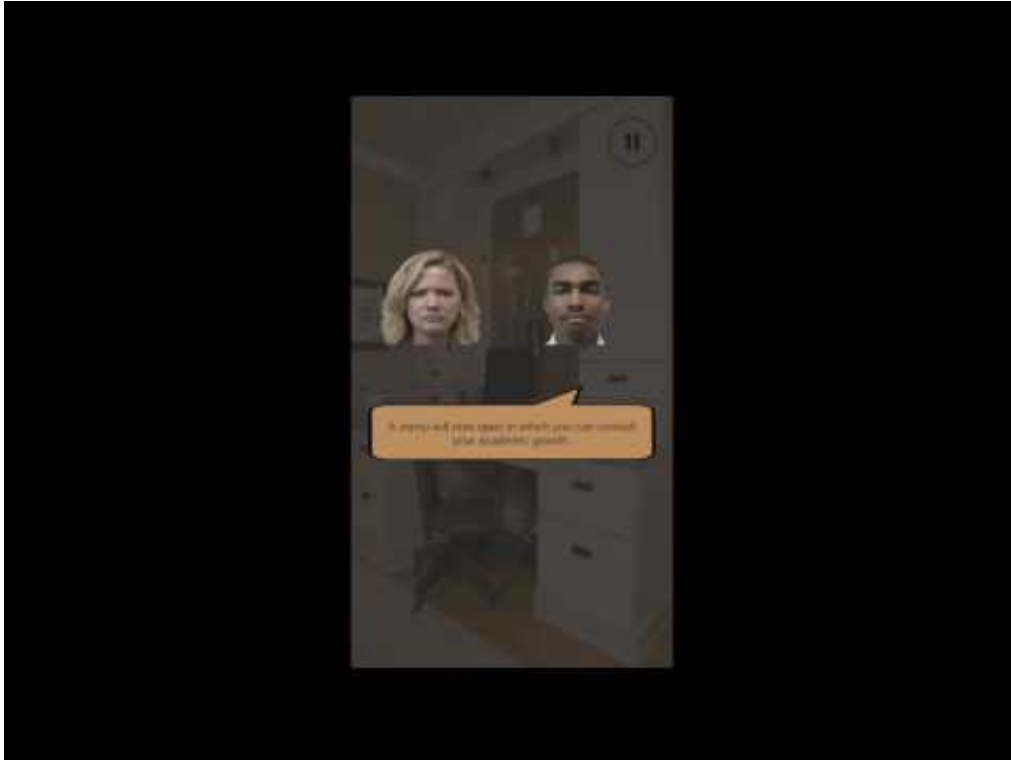
7. 7- How often do you play games that require you to interact with Non-Player Characters (virtual characters that you don't control)?

Marcar apenas uma oval.

- I avoid games where I have to interact with NPCs
- I occasionally play games where I have to interact with NPCs
- I prefer playing games where I have to interact with NPCs

Please watch the video until the end before you answer the questions. Please watch the video in fullscreen by clicking its youtube link so you can read the agents' lines. Be sure to pay attention to the Tutor's words and behavior:

Video 1



<http://youtube.com/watch?v=kbao3X1L4Vo>

8. 1- How Important was the Tutors' behavior characteristic shown on this video for a Virtual Tutor?

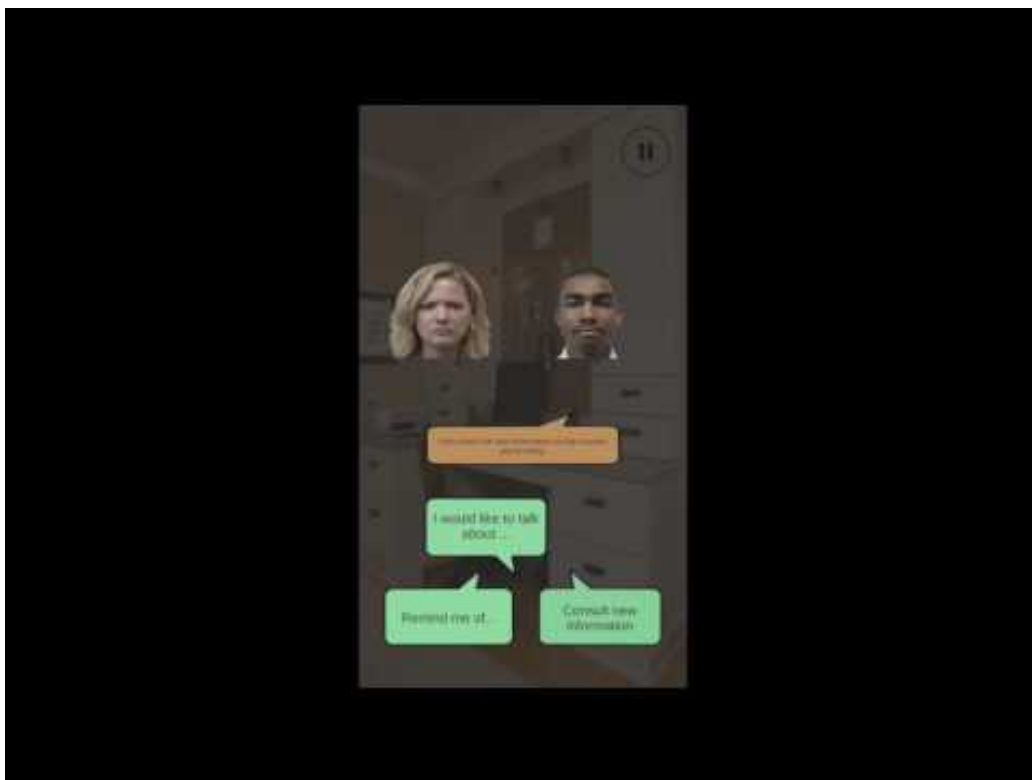
Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Not Important at all	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Fundamental

9. 2- Describe the Tutors' behavior in this video using up to three adjectives *

Please watch the video until the end before you answer the questions. Please watch the video in fullscreen by clicking its youtube link so you can read the agents' lines. Be sure to pay attention to the Tutor's words and behavior:

Video 2



<http://youtube.com/watch?v=xJwyuYZt63Q>

10. 1- How Important was the Tutors' behavior characteristic shown on this video for a Virtual Tutor? *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Not Important at all	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Fundamental

11. 2- Describe the Tutors' behavior in this video using up to three adjectives *

Please watch the video until the end before you answer the questions. Please watch the video in fullscreen by clicking its youtube link so you can read the agents' lines. Be sure to pay attention to the Tutor's words and behavior:

Video 3



<http://youtube.com/watch?v=G25ggIW1vD4>

12. 1- How Important was the Tutors' behavior characteristic shown on this video for a Virtual Tutor? *

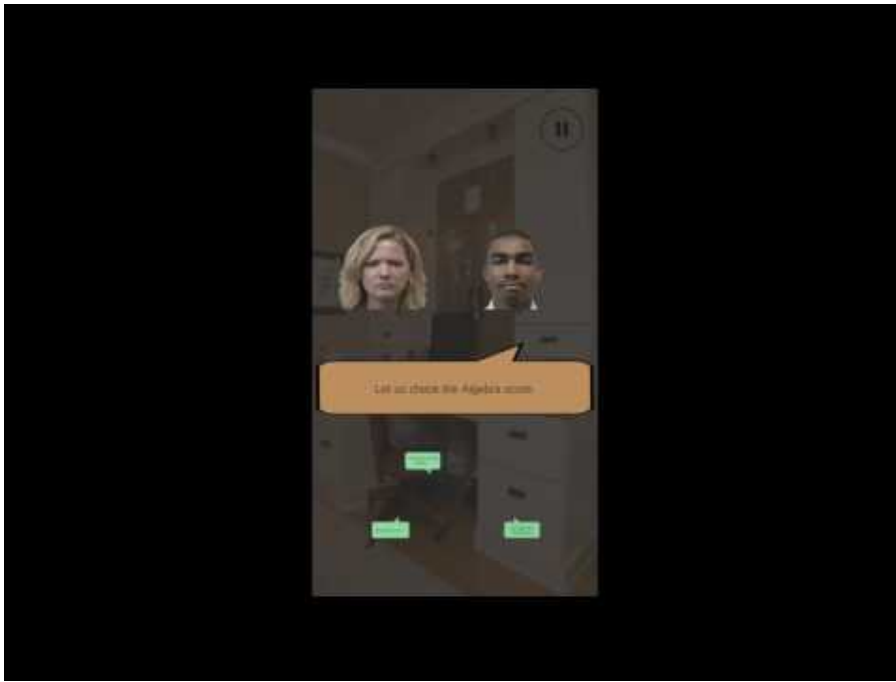
Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Not Important at all	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Fundamental

13. 2- Describe the Tutors' behavior in this video using up to three adjectives *

Please watch the video until the end before you answer the questions. Please watch the video in fullscreen by clicking its youtube link so you can read the agents' lines. Be sure to pay attention to the Tutor's words and behavior:

Video 4



<http://youtube.com/watch?v=ETvgw2-NsG0>

14. 1- How Important was the Tutors' behavior characteristic shown on this video for a Virtual Tutor? *

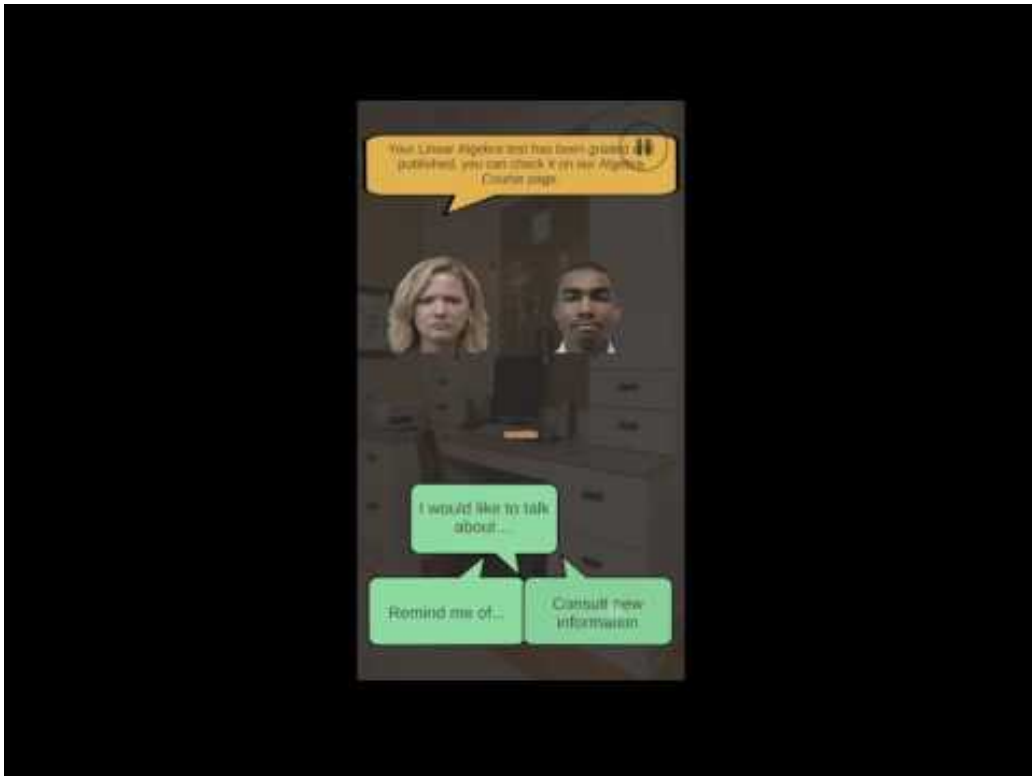
Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Not Important at all	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Fundamental

15. 2- Describe the Tutors' behavior in this video using up to three adjectives *

Please watch the video until the end before you answer the questions. Please watch the video in fullscreen by clicking its youtube link so you can read the agents' lines. Be sure to pay attention to the Tutor's words and behavior:

Video 5



<http://youtube.com/watch?v=ULvfucRIDg>

16. 1- How Important was the Tutors' behavior characteristic shown on this video for a Virtual Tutor? *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Not Important at all	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Fundamental

17. 2- Describe the Tutors' behavior in this video using up to three adjectives *

Please watch the video until the end before you answer the questions. Please watch the video in fullscreen by clicking its youtube link so you can read the agents' lines. Be sure to pay attention to the Tutor's words and behavior:

Video 6



<http://youtube.com/watch?v=WBFDDe-3yDk>

18. 1- How Important was the Tutors' behavior characteristic shown on this video for a Virtual Tutor? *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Not Important at all	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Fundamental

19. 2- Describe the Tutors' behavior in this video using up to three adjectives *

B

Demographic statistic results

Statistics

Age		
N	Valid	30
	Missing	0
Mean		23,60
Median		23,50
Mode		24
Std. Deviation		4,484

Figure B.1: Participants' Age Statistics

Gender					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	18	60,0	60,0	60,0
	Male	12	40,0	40,0	100,0
Total		30	100,0	100,0	

Figure B.2: Participants' gender statistics



Figure B.3: Pie chart with the participants' gender

		Student			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	University	18	60,0	60,0	60,0
	Not a student	7	23,3	23,3	83,3
	Student out of university	5	16,7	16,7	100,0
Total		30	100,0	100,0	

Figure B.4: Participants' student information statistics

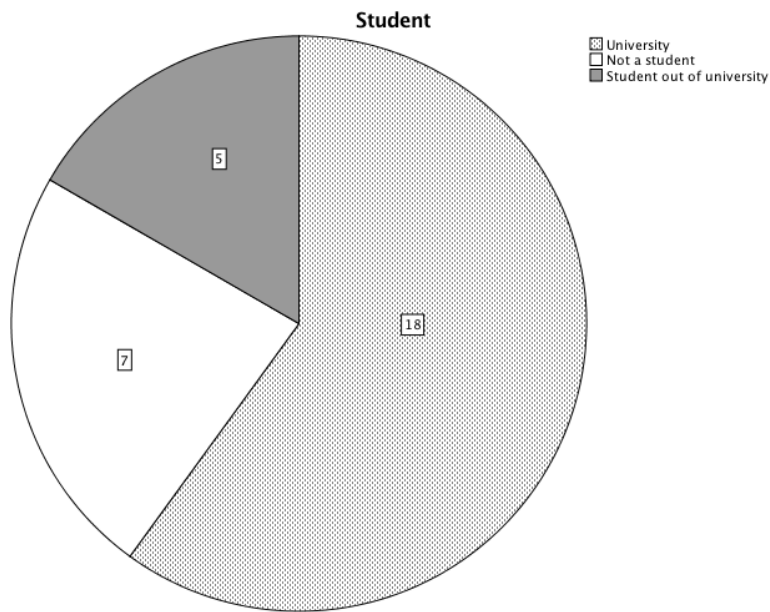


Figure B.5: Pie chart with the participants' student information

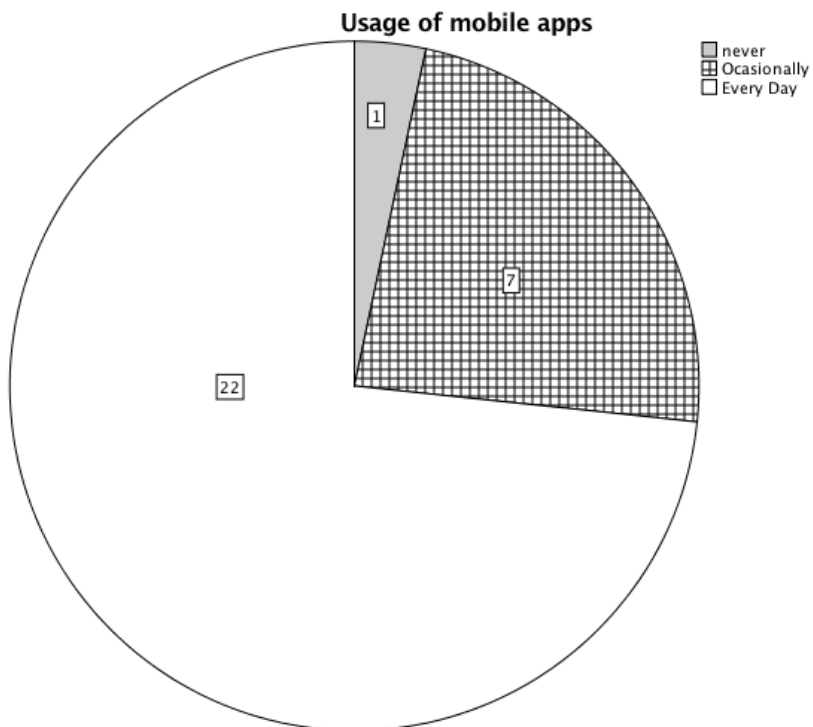


Figure B.6: Pie chart with the participants' information on the usage of mobile apps on their smartphone

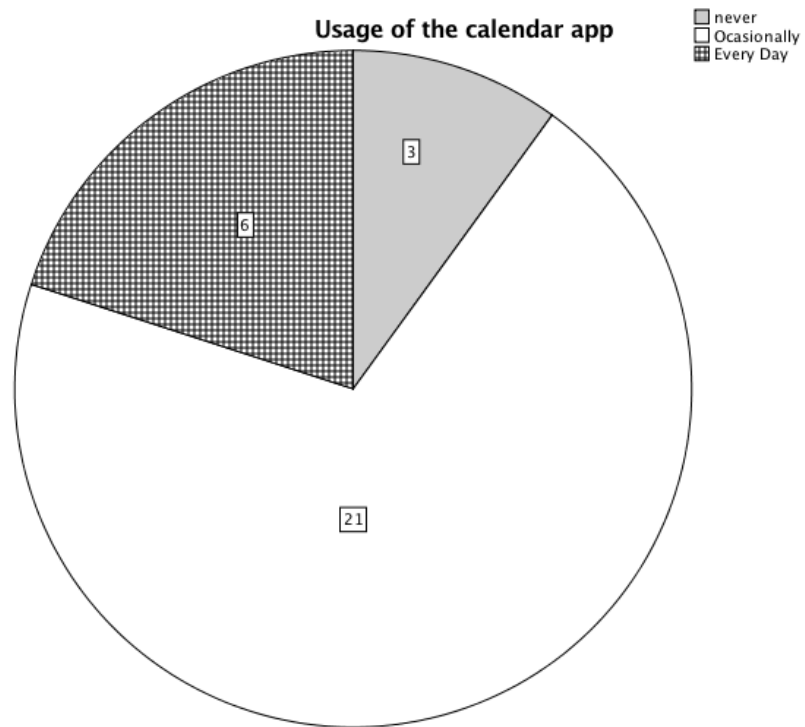


Figure B.7: Pie chart with the participants' information on the usage of the calendar apps on their smartphone

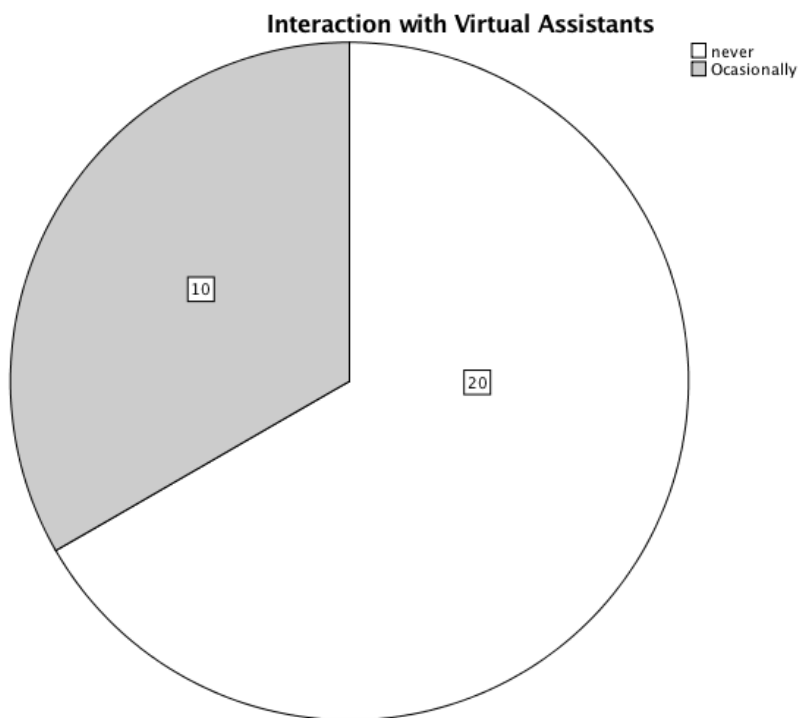


Figure B.8: Pie chart with the participants' information on the familiarity with Virtual Assistants

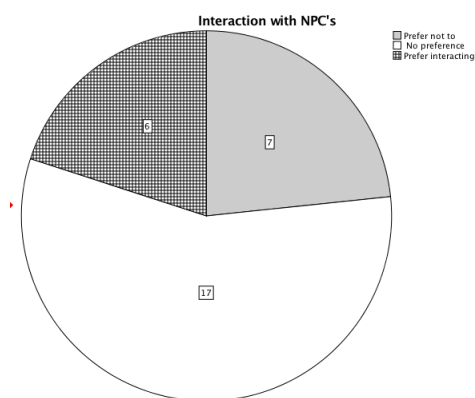
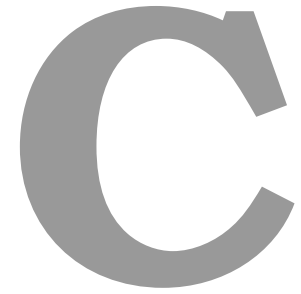


Figure B.9: Pie chart with the participants' information on the preference to interact with NPC's in games



Lickert Scales

Acquaintance_Importance

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not important at all	1	3,2	3,3	3,3
	2	2	6,5	6,7	10,0
	3	1	3,2	3,3	13,3
	4	1	3,2	3,3	16,7
	5	5	16,1	16,7	33,3
	6	5	16,1	16,7	50,0
	7	6	19,4	20,0	70,0
	8	6	19,4	20,0	90,0
	Fundamental	3	9,7	10,0	100,0
	Total	30	96,8	100,0	
Missing	System	1	3,2		
Total		31	100,0		

Figure C.1: Frequency table regarding the importance of the video focusing on the acquaintance level of the relationship

Maintenance_Importance

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not important at all	2	6,5	6,7	6,7
	2	1	3,2	3,3	10,0
	3	4	12,9	13,3	23,3
	5	4	12,9	13,3	36,7
	6	3	9,7	10,0	46,7
	7	12	38,7	40,0	86,7
	8	1	3,2	3,3	90,0
	Fundamental	3	9,7	10,0	100,0
	Total	30	96,8	100,0	
	Missing	System	1	3,2	
Total		31	100,0		

Figure C.2: Frequency table regarding the importance of the video focusing on the maintenance level of the relationship

SelfDisc_Importance

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not important at all	1	3,2	3,3	3,3
	3	2	6,5	6,7	10,0
	4	6	19,4	20,0	30,0
	5	7	22,6	23,3	53,3
	6	3	9,7	10,0	63,3
	7	7	22,6	23,3	86,7
	8	3	9,7	10,0	96,7
	Fundamental	1	3,2	3,3	100,0
	Total	30	96,8	100,0	
	Missing	System	1	3,2	
Total		31	100,0		

Figure C.3: Frequency table regarding the importance of the video focusing on the rapport strategy of self disclosure for positivity

BackChannel_Importance

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3	4	12,9	13,3	13,3
	4	2	6,5	6,7	20,0
	5	4	12,9	13,3	33,3
	6	5	16,1	16,7	50,0
	7	6	19,4	20,0	70,0
	8	6	19,4	20,0	90,0
	Fundamental	3	9,7	10,0	100,0
	Total	30	96,8	100,0	
Missing	System	1	3,2		
Total		31	100,0		

Figure C.4: Frequency table regarding the importance of the video focusing on the rapport strategy of back channeling for coordination and mutual attention

Mimicry_Importance

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2	1	3,2	3,3	3,3
	3	6	19,4	20,0	23,3
	4	3	9,7	10,0	33,3
	5	4	12,9	13,3	46,7
	6	3	9,7	10,0	56,7
	7	6	19,4	20,0	76,7
	8	6	19,4	20,0	96,7
	Fundamental	1	3,2	3,3	100,0
	Total	30	96,8	100,0	
Missing	System	1	3,2		
Total		31	100,0		

Figure C.5: Frequency table regarding the importance of the video focusing on the empathy strategy of posture mimicry

Perspective_Importance

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Important at all	3	9,7	10,0	10,0
	3	2	6,5	6,7	16,7
	4	1	3,2	3,3	20,0
	5	2	6,5	6,7	26,7
	6	6	19,4	20,0	46,7
	7	5	16,1	16,7	63,3
	8	8	25,8	26,7	90,0
	Fundamental	3	9,7	10,0	100,0
	Total	30	96,8	100,0	
	Missing	System	1	3,2	
Total		31	100,0		

Figure C.6: Frequency table regarding the importance of the video focusing on the empathy strategy of perspective taking

D

Adjectivation statistic results

		Statistics					
		Acquaintance Positive	Maintenance Positive	SelfDiscPositive	BackChannel Positive	MimicryPositive	PerspectivePositive
N	Valid	30	30	30	30	30	30
	Missing	1	1	1	1	1	1
Mean		1,63	1,70	1,37	1,43	1,43	1,77
Mode		1	3	1	1	0	1

Figure D.1: Statistic information on the number positive adjectives used to describe each tutor

		Statistics					
		Acquaintance Negative	Maintenance Negative	SelfDiscNegative	BackChannel Negative	MimicryNegative	PerspectiveNegative
N	Valid	30	30	30	30	30	30
	Missing	1	1	1	1	1	1
Mean		,60	,40	,77	,53	,70	,23
Mode		0	0	0	0	0	0

Figure D.2: Statistic information on the number negative adjectives used to describe each tutor

		Statistics					
		NumberEmotionalAcquaintance	NumberEmotionalMaintenance	NumberEmotionalSelfDisc	NumberEmotionalBackChannel	NumberEmotionalMimicry	NumberEmotionalPerspective
N	Valid	30	30	30	30	30	30
	Missing	1	1	1	1	1	1
Mean		,20	,30	,47	,30	,50	,53
Mode		0	0	0	0	0	0

Figure D.3: Statistic information on the number times each tutor was qualified as showing emotions

		Statistics					
		NumberSeriousAcquaintance	NumberSeriousMaintenance	NumberSeriousSelfDisc	NumberSeriousBackChannel	NumberSeriousMimicry	NumberSeriousPerspective
N	Valid	30	30	30	30	30	30
	Missing	1	1	1	1	1	1
Mean		,53	,10	,00	,17	,17	,10
Mode		0	0	0	0	0	0

Figure D.4: Statistic information on the number times each tutor was qualified as being serious

		Statistics					
		NumberFriendlyAcquaintance	NumberFriendlyMaintenance	NumberFriendlySelfDisclosure	NumberFriendlyBackChannel	NumberFriendlyMimicry	NumberFriendlyPerspective
N	Valid	30	30	30	30	30	30
	Missing	1	1	1	1	1	1
Mean		,13	,57	,33	,27	,10	,10
Mode		0	0	0	0	0	0

Figure D.5: Statistic information on the number times each tutor was qualified as being friendly

		Statistics					
		SuppHelpAcquaintance	SuppHelpMaintenance	SuppHelpSelfDisc	SuppHelpBackChannel	SuppHelpMimicry	SuppHelpPerspective
N	Valid	30	30	30	30	30	30
	Missing	1	1	1	1	1	1
Mean		,50	,67	,23	,33	,47	,67
Mode		0 ^a	0	0	0	0	1

a. Multiple modes exist. The smallest value is shown

Figure D.6: Statistic information on the number times each tutor was qualified as being supportive or helpful



Shapiro-Wilk and Freidman test results

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Acquaintance_Importance	,158	30	,054	,920	30	,027

a. Lilliefors Significance Correction

Figure E.1: Shapiro-Wilk's results on the importance of the acquaintance phase of the relationship in the Tutors

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Maintenance_Importance	,241	30	,000	,888	30	,004

a. Lilliefors Significance Correction

Figure E.2: Shapiro-Wilk's results on the importance of the maintenance phase of the relationship in the Tutors

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
SelfDisc_Importance	,157	30	,056	,956	30	,246

a. Lilliefors Significance Correction

Figure E.3: Shapiro-Wilk's results on the importance of the rapport strategy of self disclosure for the increase of positivity in the Tutors

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
BackChannel_Importance	,158	30	,055	,925	30	,036

a. Lilliefors Significance Correction

Figure E.4: Shapiro-Wilk's results on the importance of the rapport strategy of back channeling for the increase of coordination and mutual attention in the Tutors

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Mimicry_Importance	,182	30	,013	,916	30	,021

a. Lilliefors Significance Correction

Figure E.5: Shapiro-Wilk's results on the importance of the empathy generation strategy of posture mimicry in the Tutors

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Perspective_Importance	,205	30	,002	,867	30	,001

a. Lilliefors Significance Correction

Figure E.6: Shapiro-Wilk's results on the importance of the empathy generation strategy of perspective taking in the Tutors

Friedman Test

Ranks

	Mean Rank
Acquaintance_Importance	3,58
Maintenance_Importance	3,23
SelfDisc_Importance	3,20
BackChannel_Importance	3,72
Mimicry_Importance	3,28
Perspective_Importance	3,98

Test Statistics^a

N	30
Chi-Square	5,126
df	5
Asymp. Sig.	,401

a. Friedman Test

Figure E.7: Friedman test results on the variance of importances of all behaviors shown in the videos



Wilcoxon Signed rank test

	Wilcoxon signed rank tests					
	Acquaintance	Maintenance	Self-Disclosure	Back-Channeling	Mimicry	Perspective Taking
Acquaintance	X	0.513	0.2	0.882	0.409	0.706
Maintenance	X	X	0.597	0.452	0.859	0.316
Self-Disclosure	X	X	X	0.055	0.782	0.136
Back-Channeling	X	X	X	X	0.180	0.818
Mimicry	X	X	X	X	X	0.238
Perspective Taking	X	X	X	X	X	X

Figure F.1: Matrix of significance of all the pairs of Wilcoxon signed rank tests performed



Tables from Split Database Friedman Tests

Friedman Test

Ranks

	Mean Rank
Acquaintance_Importance	3,58
Maintenance_Importance	3,38
SelfDisc_Importance	3,08
BackChannel_Importance	3,79
Mimicry_Importance	3,08
Perspective_Importance	4,08

Test Statistics^a

N	12
Chi-Square	3,095
df	5
Asymp. Sig.	,685

a. Friedman Test

Figure G.1: Results of the Friedman test counting only participants of gender male

Friedman Test

Ranks

	Mean Rank
Acquaintance_Importance	3,58
Maintenance_Importance	3,14
SelfDisc_Importance	3,28
BackChannel_Importance	3,67
Mimicry_Importance	3,42
Perspective_Importance	3,92

Test Statistics^a

N	18
Chi-Square	2,560
df	5
Asymp. Sig.	,767

a. Friedman Test

Figure G.2: Results of the Friedman test counting only participants of gender male

Friedman Test

Ranks

	Mean Rank
Acquaintance_Importance	3,75
Maintenance_Importance	3,25
SelfDisc_Importance	3,00
BackChannel_Importance	3,50
Mimicry_Importance	4,00
Perspective_Importance	3,50

Test Statistics^a

N	6
Chi-Square	1,278
df	5
Asymp. Sig.	,937

a. Friedman Test

Figure G.3: Results of the Friedman test counting only participants of age 24

Friedman Test

Ranks

	Mean Rank
Acquaintance_Importance	3,06
Maintenance_Importance	3,11
SelfDisc_Importance	3,00
BackChannel_Importance	3,78
Mimicry_Importance	3,56
Perspective_Importance	4,50

Test Statistics^a

N	9
Chi-Square	4,755
df	5
Asymp. Sig.	,446

a. Friedman Test

Figure G.4: Results of the Friedman test counting only participants of age higher than 24

Friedman Test

Ranks

	Mean Rank
Acquaintance_Importance	3,83
Maintenance_Importance	3,30
SelfDisc_Importance	3,40
BackChannel_Importance	3,77
Mimicry_Importance	2,83
Perspective_Importance	3,87

Test Statistics^a

N	15
Chi-Square	4,473
df	5
Asymp. Sig.	,483

a. Friedman Test

Figure G.5: Results of the Friedman test counting only participants of age lower than 24

Friedman Test

Ranks

	Mean Rank
Acquaintance_Importance	3,25
Maintenance_Importance	3,42
SelfDisc_Importance	3,50
BackChannel_Importance	3,42
Mimicry_Importance	3,25
Perspective_Importance	4,17

Test Statistics^a

N	6
Chi-Square	1,355
df	5
Asymp. Sig.	,929

a. Friedman Test

Figure G.6: Results of the Friedman test counting only participants that always use their calendar app

Friedman Test

Ranks

	Mean Rank
Acquaintance_Importance	4,17
Maintenance_Importance	3,50
SelfDisc_Importance	4,17
BackChannel_Importance	3,83
Mimicry_Importance	1,33
Perspective_Importance	4,00

Test Statistics^a

N	3
Chi-Square	6,772
df	5
Asymp. Sig.	,238

a. Friedman Test

Figure G.7: Results of the Friedman test counting only participants that never use their calendar app

• Friedman Test

Ranks

	Mean Rank
Acquaintance_Importance	3,60
Maintenance_Importance	3,14
SelfDisc_Importance	2,98
BackChannel_Importance	3,79
Mimicry_Importance	3,57
Perspective_Importance	3,93

Test Statistics^a

N	21
Chi-Square	4,725
df	5
Asymp. Sig.	,450

a. Friedman Test

Figure G.8: Results of the Friedman test counting only participants that occasionally use their calendar app

Friedman Test

Ranks

	Mean Rank
Acquaintance_Importance	3,33
Maintenance_Importance	2,33
SelfDisc_Importance	2,92
BackChannel_Importance	3,92
Mimicry_Importance	4,50
Perspective_Importance	4,00

Test Statistics^a

N	6
Chi-Square	6,412
df	5
Asymp. Sig.	,268

a. Friedman Test

Figure G.9: Results of the Friedman test counting only participants that prefer games where they have to interact with a NPC

Friedman Test

Ranks

	Mean Rank
Acquaintance_Importance	3,64
Maintenance_Importance	3,57
SelfDisc_Importance	4,29
BackChannel_Importance	4,14
Mimicry_Importance	2,14
Perspective_Importance	3,21

Test Statistics^a

N	7
Chi-Square	8,295
df	5
Asymp. Sig.	,141

a. Friedman Test

Figure G.10: Results of the Friedman test counting only participants that prefer games where they are not forced to interact with NPC

Friedman Test

Ranks

	Mean Rank
Acquaintance_Importance	3,65
Maintenance_Importance	3,41
SelfDisc_Importance	2,85
BackChannel_Importance	3,47
Mimicry_Importance	3,32
Perspective_Importance	4,29

Test Statistics^a

N	17
Chi-Square	6,209
df	5
Asymp. Sig.	,286

a. Friedman Test

Figure G.11: Results of the Friedman test counting only participants that have no preferences about the presence of NPCs in a game

Friedman Test

Ranks

	Mean Rank
Acquaintance_Importance	3,57
Maintenance_Importance	3,05
SelfDisc_Importance	3,14
BackChannel_Importance	3,86
Mimicry_Importance	3,55
Perspective_Importance	3,84

Test Statistics^a

N	22
Chi-Square	4,403
df	5
Asymp. Sig.	,493

a. Friedman Test

Figure G.12: Results of the Friedman test counting only participants that use mobile apps everyday

Friedman Test

Ranks

	Mean Rank
Acquaintance_Importance	3,63
Maintenance_Importance	3,75
SelfDisc_Importance	3,38
BackChannel_Importance	3,31
Mimicry_Importance	2,56
Perspective_Importance	4,38

Test Statistics^a

N	8
Chi-Square	5,230
df	5
Asymp. Sig.	,388

a. Friedman Test

Figure G.13: Results of the Friedman test counting only participants that do not use mobile apps everyday

Friedman Test

Ranks

	Mean Rank
Acquaintance_Importance	3,50
Maintenance_Importance	3,17
SelfDisc_Importance	3,36
BackChannel_Importance	4,14
Mimicry_Importance	3,03
Perspective_Importance	3,81

Test Statistics^a

N	18
Chi-Square	5,358
df	5
Asymp. Sig.	,374

a. Friedman Test

Figure G.14: Results of the Friedman test counting only participants that are university students

Friedman Test

Ranks

	Mean Rank
Acquaintance_Importance	3,71
Maintenance_Importance	3,33
SelfDisc_Importance	2,96
BackChannel_Importance	3,08
Mimicry_Importance	3,67
Perspective_Importance	4,25

Test Statistics^a

N	12
Chi-Square	4,603
df	5
Asymp. Sig.	,466

a. Friedman Test

Figure G.15: Results of the Friedman test counting only participants that are not university students

Friedman Test

Ranks

	Mean Rank
Acquaintance_Importance	3,55
Maintenance_Importance	3,35
SelfDisc_Importance	3,23
BackChannel_Importance	3,85
Mimicry_Importance	3,23
Perspective_Importance	3,80

Test Statistics^a

N	20
Chi-Square	2,558
df	5
Asymp. Sig.	,768

a. Friedman Test

Figure G.16: Results of the Friedman test counting only participants that never had interacted with a Virtual Assistant

Friedman Test

Ranks

	Mean Rank
Acquaintance_Importance	3,65
Maintenance_Importance	3,00
SelfDisc_Importance	3,15
BackChannel_Importance	3,45
Mimicry_Importance	3,40
Perspective_Importance	4,35

Test Statistics^a

N	10
Chi-Square	4,313
df	5
Asymp. Sig.	,505

a. Friedman Test

Figure G.17: Results of the Friedman test counting only participants that had occasionally interacted with a Virtual Assistant