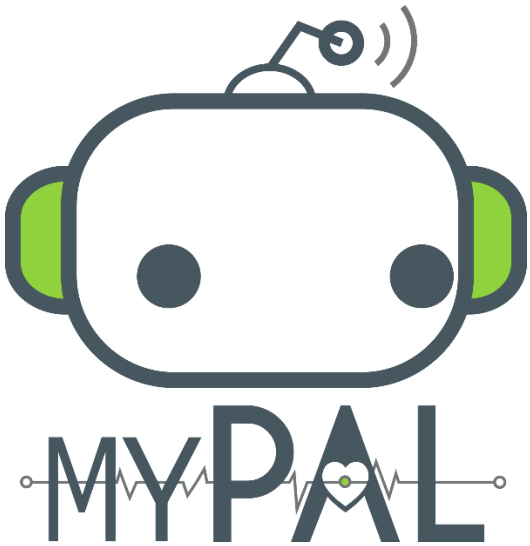
Master Thesis Artificial Intelligence



MyPAL: A Digital Diabetes Diary with a Responsive Social Avatar

Mike Ligthart, BSc

Radboud University Nijmegen

June 2016

Internal Supervisor: dr. Pim Haselager
Radboud University Nijmegen

External Supervisor: prof. dr. Mark Neerincx
TNO and Delft University of Technology

Reviewer: dr. Khiet Truong
University of Twente

Radboud University

TNO innovation for life

Abstract

Diabetes Mellitus type I is an incurable disease that can be diagnosed at a young age. A structured lifestyle, where insulin use, carbohydrate intake and blood glucose are regularly monitored, is the only path to a relatively normal life. Children and their parents must remain vigilant. This lifestyle is especially demanding for children and not every child is as good in self-management as they need to be. They can use some help with this.

MyPAL is a digital diabetes diary that children can use to record their insulin use, carbohydrate intake and blood glucose values as well as write something about their day and how they feel. With that information the children can more easily link their diabetes values, what they eat and how they feel together. With this insight they can manage their insulin use and diet more efficiently. Besides children also medical professionals, parents and researchers benefit from this information. For example, a diabetes nurse can improve the treatment plan, parents can get a better idea how their child is doing and researchers can investigate the relationship between food, mood and blood glucose values more closely.

The only constraint is that the required information is added regularly. And if children find something difficult it is consistently keeping a diary. One of the most mentioned causes of why children have trouble to consistency use a diabetes diary is the lack of motivation. MyPAL provides several solutions for this problem. Following the situated Cognitive Engineering (sCE) design method, myPAL is specifically designed to support the development of motivation.

The first deliverable included in this thesis are the system specifications of myPAL. These specifications are based on operational demands, human-factor knowledge, and technological principles. The self-determination theory, that identified the feeling of autonomy, competence and relatedness as antecedents of motivation, is the largest human-factor contribution. An avatar, a technological principle, is used to support those antecedents of motivation. By autonomously responding to the added content in a social fashion, e.g. matching the gestures and speech of the avatar appropriately to the mood of the child, children feel more supported in their competence and relatedness.

The second deliverable is a thorough evaluation of the avatar behaviors and its effects on the attitude of the children towards the robot, motivation support and performance. Performance is measured in terms of the amount and the consistency of the added content. A three-week user study with 13 children with diabetes was performed for this evaluation. Results show that almost all the avatar behaviors are picked up by the children and that those behaviors positively affect the motivation and performance of the children.

The final deliverables are two design recommendations that have been found to modulate the effectiveness of myPAL. The first is avatar *quality over quantity*. The avatar behavior must be appreciated by the children in order to be effective. Simply showing to avatar more does not increase the appreciation. The behavior must match the children's preferences. The second recommendation is *avatar sociability is key*. The more social the behavior of the avatar is, the more it is appreciated by the child and the more motivated the children are to add more content consistently.

The system specifications, evaluation and design recommendations bring the sustainable application of autonomous avatars in diabetes care a step closer to being realized.

Abstract (Dutch)

Diabetes Mellitus type I is een ongeneselijke ziekte dat zich al op jonge leeftijd kan openbaren. Een gestructureerde dagindeling, waarbij insulinegebruik, inname van koolhydraten en de bloedglucosewaarden regelmatig worden bijgehouden, is de enige weg richting een relatief normaal leven. Kinderen en hun ouders dienen voortdurend waakzaam te zijn. Deze manier van leven is voor kinderen in het bijzonder erg veeleisend. Niet elk kind is zelfredzaam genoeg. Ze kunnen hulp gebruiken om dat te worden.

MyPAL is een digitaal diabetesdagboek dat kinderen kunnen gebruiken om hun insulinegebruik, koolhydrateninname en bloedglucosewaarden bij te houden. Verder kunnen ze hun dagelijkse bezigheden en gevoel hierbij noteren. Met die informatie kunnen kinderen het verband leggen tussen hun diabeteswaarden, wat ze eten en hoe ze zich voelen. Met dat inzicht kunnen ze hun insuline inname en dieet beter reguleren. Naast kinderen kunnen ook medische professionals, ouders en onderzoekers profiteren van de data. Een diabetesverpleegkundige kan bijvoorbeeld het behandelplan bijstellen, ouders krijgen meer inzicht in hoe het gaat met hun kind en onderzoekers kunnen de relatie tussen voedsel, gevoel en diabeteswaarden nog gedetailleerder onderzoeken.

Een voorwaarde is echter wel dat de vereiste informatie regelmatig wordt toegevoegd aan het dagboek. Als kinderen ergens moeite mee hebben dan is het een dagboek consistent bijhouden. Eén van de meest genoemde redenen hiervoor is dat kinderen nauwelijks gemotiveerd zijn. MyPAL biedt meerdere oplossingen voor dit probleem. Met behulp van de 'situated Cognitive Engineering' ontwerpmethode is myPAL specifiek ontworpen om de motivatieontwikkeling bij kinderen te ondersteunen.

Het eerste wat deze scriptie oplevert zijn de systeemspecificaties voor myPAL. De specificaties zijn gefundeerd op de functionele eisen, human-factors kennis en technologische principes. De zelfbeschikkingstheorie, wat het gevoel van autonomie, competentie en verbondenheid heeft geïdentificeerd als voorlopers van motivatie, wordt ingebracht vanuit de human-factors kennis. Een avatar, als technologisch principe, kan vervolgens ingezet worden om die voorlopers van motivatie te ondersteunen. Door de avatar autonoom en op sociale wijze te laten reageren op de toegevoegde dagboekinhoud, bijvoorbeeld door de bewegingen en de spraak van de avatar af te stemmen op de gemoedsruststand van het kind, voelen kinderen zich competenter en meer verbonden.

Het tweede wat deze scriptie oplevert is een grondige evaluatie van het gedrag van de avatar en zijn effecten op de attitude van de kinderen ten opzichte van de avatar, ontwikkeling van motivatie en de prestatie van de kinderen. Een experiment van drie weken met dertien kinderen met diabetes is uitgevoerd ten behoeve van de evaluatie. De resultaten tonen aan dat het meeste van de avatargedragingen worden opgepikt door de kinderen. Belangrijker nog, de opgepikte gedragingen hebben een positief effect op de motivatie en prestatie van de kinderen.

Het laatste wat deze scriptie oplevert zijn twee aanbevelingen die van invloed zijn op de effectiviteit van myPAL. De eerste is *kwaliteit boven kwantiteit*. Het avatargedrag moet gewaardeerd worden door de kinderen om effectief te kunnen zijn. Simpelweg het laten zien van de avatar maakt het niet meer gewaardeerd. Het gedrag moet passen bij de voorkeuren van de kinderen. De tweede

aanbeveling is: *het sociale vermogen van de avatar is de sleutel tot succes.* Des te sociale het gedrag van de avatar, des te meer het wordt gewaardeerd door de kinderen en des te meer de kinderen gemotiveerd zijn om het dagboek consistent van inhoud te voorzien.

De systeemspecificaties, de grondige evaluatie en de ontwerpsuggesties brengen de duurzame inzet van autonome avatars in de diabeteszorg een stap dichter bij de werkelijkheid.

Table of contents

1. Introduction

More than half a million children suffer from Diabetes Mellitus type I worldwide. This number increases every year. A quarter of the children with Diabetes live in Europe. Children benefit from a structured lifestyle where insulin use, carbohydrate intake, and blood glucose are regularly monitored. Starting at a young age with a well-balanced self-management plan including self-reports, physical activity, and a healthy diet increases the changes for a relatively normal life [1].

Helping children with diabetes to increase their self-management skills is the path to a better quality of life. This is the main force that drives the Horizon2020 project: Personal Assistant for a healthy Lifestyle (PAL). The `personal assistant' comes in different forms depending on the user, location and context. For example, for children visiting the hospital the personal assistant is a robot while at home it is a virtual version of that robot. At the hospital the children learn more about diabetes together with the robot in a playful fashion. At home children can use various mobile health applications. In those apps the robot is virtually present. Besides children also their parents and medical professionals can use the PAL-system. The project is carried out by a European consortium with partners from The Netherlands, Italy, Germany and the United Kingdom. Research institute TNO has the role of project coordinator [2].

My master graduation project is a part of the PAL project. During a 10-month internship at TNO Soesterberg I developed one of the mobile health apps that will be part of PAL, namely a digital diabetes diary called myPAL. Research shows that keeping a digital diabetes diary, including for example insulin use and carbohydrate intake, result in more insight and empowerment for the patient [3], [4]. Furthermore, it enables caregivers to make more informed adaptations to the treatment plan [5]. For adults dozens of mobile apps are available to keep a diabetes diary [6] but for children different design choices are necessary. For example, motivating children to keep a diary on a regular basis is challenging. Coercing a child often is counterproductive [7].

In the first half of this thesis I describe the groundwork that is required to design, build and evaluate myPAL: a digital diabetes diary with a responsive avatar. The goal of myPAL is to support the motivation of children to add content to the diary as consistently as possible. Human-factor knowledge and technological principles come together to achieve this goal. The two main building blocks are the self-determination theory, which is a framework for supporting intrinsic motivation, and a social responsive avatar. Using rapid-prototyping techniques and a larger pilot study a suitable configuration of all the building blocks is created in the form of system design specifications and a prototype. The design process mainly revolved around finding suitable ways of supporting motivation with myPAL.

If previous research shows us anything it is that children respond differently to systems like myPAL [5]. Responsive social systems vary in effectiveness and are not appreciated by everybody equally. The keyword for tackling these problems is personalization. If responsive systems were able to adapt their behavior to a specific user a higher effectiveness and appreciation is to be expected. The thing is, we are not there yet. More insight is needed into how children respond to a responsive social system such as myPAL.

In the second half of this thesis I describe a three-week user study of children using myPAL. I monitored the development of motivation and the performance over time. Performance is defined

by the amount and the consistency of the content added to the diary. The first aim for this user study is to evaluate the design. Does the implemented avatar behavior get picked up by the children? Does the avatar behavior influence motivation and performance? The second aim of the user study is to achieve more insight in how children respond to myPAL in terms of performance and motivational development. To do this I will categorize the children based on their motivation and performance. Can we subsequently identify predictors that are able to indicate whether children are currently on a path for a high performance or not?

This thesis is structured as follows. In the next chapter fundamental concepts relevant for this thesis are introduced in more detail. First, the impact of diabetes on children is introduced and I motivate why this research can have an impact on their life's. Secondly, the position of myPAL in the PAL project is discussed. In the third section a definition is given of social robots in the health care domain. Because we are doing research with children it is important to design the prototype and the user study accordingly. In the fourth background section I reflect on several ethical implications of doing research with children. In the final background section, I introduce, situated Cognitive Engineering (sCE), the user-centered design method I used to develop myPAL.

A technical solution is never stands alone; it is always embedded in a certain context. The context provides operational demands for the solution. An important principle of user-centered design is to really take the perspective of the user. Therefore, besides technological principles also human factor knowledge is included in the foundation of the solution. The foundation, including design and research questions, is further discussed in chapter 3.

The design cycle of sCE continues with extracting specifications for the system from the foundation. These specifications, discussed in chapter 4, lead to a prototype which is described in chapter 5. The pilot study and the three-week user study are discussed in chapter 6. The formulated design and research questions are also evaluated in chapter 6. Finally, the whole process as well as the overall results are evaluated in chapter 7.

2. Background

Diabetes mellitus is a serious and incurable disease. Approximately 1 in 11 people (415 million people worldwide) have diabetes. 26% of those patients do not reach the age of 60. Every 6 seconds someone dies of complications caused by diabetes [1].

Diabetes not only affects adults but also children. To deal properly with diabetes they have to pay attention constantly to what they do and what they eat. The burden of constant vigilance is not carried as well by everyone. This affects the quality of life of many children and families. The personal assistant for a healthy lifestyle (PAL) aims to be a tool for children, parents and medical professionals to help them deal with diabetes better and increase the quality of life of the young patients.

The research described in this thesis focusses on a component of the PAL system called myPAL, a digital diabetes diary with a responsive avatar. The development and evaluation process of myPAL is structured by a user-centered design method called situated Cognitive Engineering. An important aspect of that method is to first ground yourself in the domains you will be working in. The purpose of this background chapter is exactly that.

In this chapter I will give a more thorough introduction of the problem domain, children with diabetes, and the intended solution, the PAL project and social robots. Doing research with children is not something to take lightly. I will among other things address the ethical constrains of doing research with children. Last but not least, situated Cognitive Engineering also deserves a proper introduction.

2.1.Children with Diabetes Mellitus type I

Diabetes mellitus can be divided in two categories: type I and type II. Type I diabetes usually develops during childhood and is characterized by a loss of insulin-producing cells in the pancreas, leading to an insulin deficiency. Although the exact cause is unknown it is most likely caused by an auto-immune type of disease where the immune system attacks the insulin producing cells [8]. Type II diabetes is usually diagnosed in adults, although more and more children are getting diagnosed as well, and is linked to an excessive body weight and physical inactivity. It is characterized by the inability of the body to effectively use insulin. The ratio between type I and type II diabetes is about 10 – 90%. Type II is preventable in most cases by having a healthy lifestyle, unfortunately type I is not [9]. In this project we focus on children with diabetes type I.

The prevalence of type I diabetes among children is the highest in Europe. Of the 542.000 children with type I diabetes worldwide 140.000 live in Europe [1]. The insulin deficiency causes a deregulation of the blood glucose levels leading to extremes. Too much glucose in the blood, a 'hyper', can cause an increased need to urinate, extreme thirst, mood swings, lack of concentration and nausea. Too low levels of blood glucose, a 'hypo', can cause sweating, shaking, dizziness, mood swings, lack of concentration and extreme hunger. These symptoms appear almost instantly and without warming. Long-term consequences of diabetes are an increased risk for heart and cardiovascular disease and permanent damage to the eyes, kidneys and nervous system [9]. Finally, about 40% of adults suffer from psychological problems as an indirect result of their diabetes. Only 10% with mental problems receive professional help. Psychological problems not

only decrease the quality of life, it also undermines the self-management causing a downward spiral [10].

Diabetes is (in most cases) an incurable chronic illness. The only proven path to decrease the risk for complications, e.g. microvascular and neurologic damage, and a higher quality of life in general is an intensive self-managed 'diabetes therapy' [10]. Important aspects of this self-managed therapy are insulin therapy and a consistent healthy lifestyle. With insulin therapy the blood glucose levels are monitored regularly and adjusted with the help of insulin injections. This can be done manually with a diabetes pen or automatically with a diabetes pump. The pump is permanently connected to the body and measures the glucose values continuously. Most children in the Netherlands use a pump. A second important aspect of the diabetes therapy is a regular and healthy lifestyle including a custom diet omitting too much carbohydrates and enough exercise [11].

Living with diabetes can feel overwhelming for children and parents alike due to the constant vigilance they need to have for good care. It puts a great deal of stress upon families. The strict regimen of insulin therapy and dietary control can restrict the child to participate in certain social activities such as sleepovers and school field trips. As a consequence, the child may feel isolated and left out. Furthermore, usually the parents are the ones who take upon the role as care supervisor. They remind or instruct the children to complete a care task, e.g. checking the blood glucose values. Often this is experienced as nagging by the children and as disobedience by the parents leading up to tension [12]. Tension can also arise with siblings who feel left out because the parent is focusing more on the child with diabetes [13]. All these factors result in a higher risk for depression, anxiety and other psychological problems for children with diabetes [14].

It becomes evident that a good care strategy focusses on increasing the self-management skills and takes into account both the child as well as the family. When children adhere more to the diabetes regimen they not only remain healthy, they also reduce stress for themselves and for their families. It also becomes evident that children and their parents could use some help to achieve this. Help that the personal assistant for a healthy lifestyle might provide.

2.2. Personal Assistant for a healthy Lifestyle (PAL)

The personal assistant for a healthy lifestyle (PAL) is a platform that serves three main users: children with diabetes, their parents and their attending medical professionals e.g. diabetes nurse. The aim of the platform is to assist the different users in establishing a solid diabetes regime, considering the different responsibilities of each user, before adolescence. Children are assisted in developing the necessary self-management skills. Parents are guided towards an effective supportive role. Medical professionals gain access to better information to make more informed decisions and control the behavior of the platform.

The platform is a combination of several components (see Figure 1). The plans are as follows: children interact with the PAL system through a social robot, its virtual avatar and various (mobile) health applications. In the hospital children can individually interact with the PAL robot. At special diabetes summer camps multiple children can interact as a group with the PAL robot. Due to the cost children do not interact with the physical robot at home or at school but with a virtual version of the robot (the avatar). Furthermore, the children have access to several educational, medical,

personal or entertaining applications that are connected to the PAL system. For example, in the hospital they can play a sorting game on a large touch table whereas at home they can access the PAL system through a website or smartphone app.

The behavior of the robot is determined based on previous interactions, personal preference and characteristics, demographics and other important features. This behavior is determined by the core of the PAL system, a knowledge and reasoning base, situated in the cloud where all the input is stored and processed into output.

Medical professionals can control the behavior of the PAL system through an authoring and control tool. It is intended that the interaction between the PAL system and the children is semi-autonomous. Now someone for example has to explicitly select a sentence the robot is going to say. The plan is to make conversation work by itself. Instead of doing everything manually the medical professional will just have to place the robot somewhere, e.g. in front of the child, and activate the wanted functionality, e.g. play a game.

Parents will be given access to a monitoring system where they receive important information about their child, suggestions for certain actions and if needed extra help.

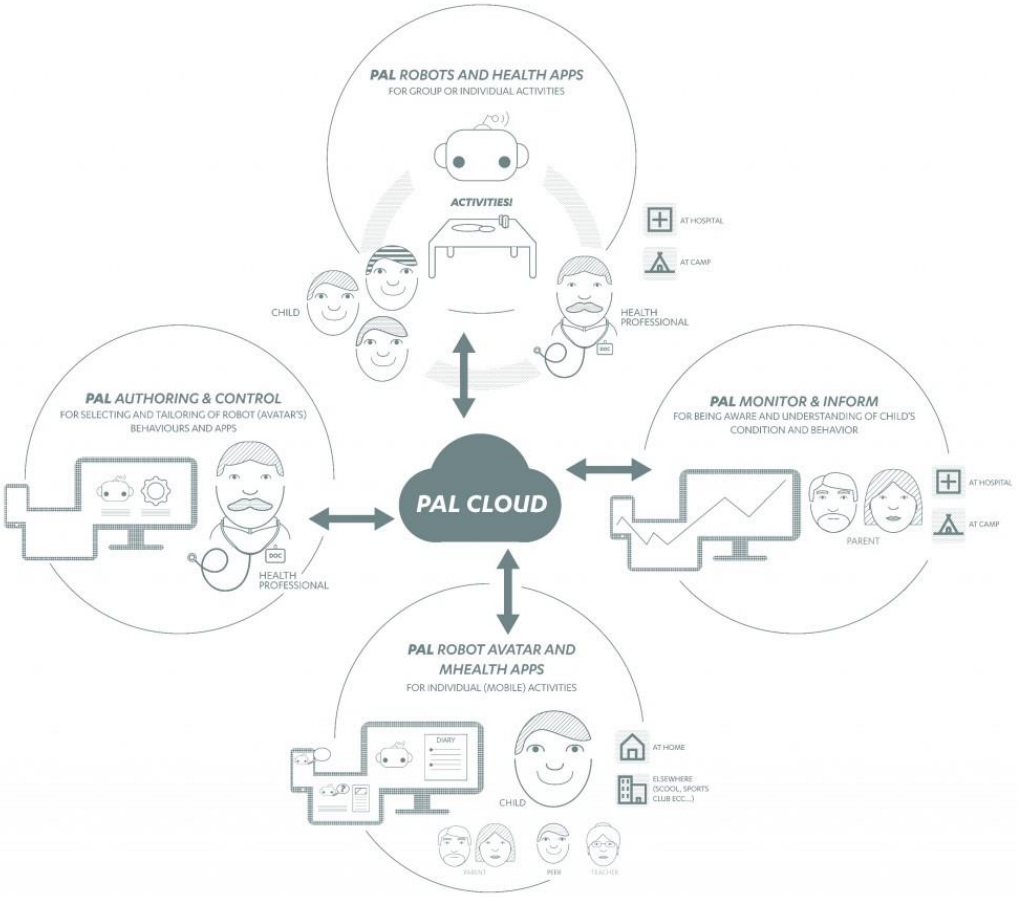


Figure 1: overview of the system of the personal assistant for a healthy lifestyle.

The PAL project is a 4-year Horizon2020 project of 4.5 million Euro's funded by the European Union. The consortium working on the PAL project consist of the research partners TNO (coordinator), Deutsches Forschungszentrum Für Künstliche Intelligenz (DFKI), Fondazione Centro San Raffaele (FCSR), Imperial Collage London and Delft University of Technology, the end-users represented by the hospitals Gelderse Vallei and Meander, and the Diabetics Associations of Netherlands and Italy, and small-medium sized enterprises (SME's) Mixel and Produxi [2]. It started in March 2015 at the same time as my internship with TNO.

It is important to keep in mind that at the start of my internship the project just started and that none of the above described was realized. However, the PAL project did not start with nothing. It could use most of the deliverables and knowledge from the ALIZ-E project which researched, among other things, whether a social robot could be used for a similar purpose in diabetes care. The result showed that a social robot is indeed an effective addition to diabetes care [5].

One of the main mobile health apps intended to be connected to the PAL system is a digital diabetes diary that includes an embedded avatar of the social robot. Previous ALIZ-E research determined that when an avatar responds in a certain way to the added diary content children were more inclined to use and like the diary more. A dialogue model was used to determine how the robot should respond. The model consisted of pre-constructed blocks of speech and gesture outputs that could be selected by the experiment leader. In was also the experiment leader who interpreted the user input (speech, gestures and actions) [5].

The challenge for my internship was to design, build and evaluate a digital diabetes diary with a responsive social avatar that operates autonomously. This means that no human operator controls of the avatar. A second challenge was the overcome the lack of available resources. There was no suitable diary that provided the necessary customizability and there was no avatar. I have built all these elements from scratch to create myPAL.

2.3.Social robots in health care

Both the physical PAL robot as its virtual copy in myPAL can be called a social robot. It is still an ongoing discussion whether a virtual agent can be called a robot. I would call any embodied artificial intelligent agent a robot. It does not matter for the definition whether the instantiation of the body is physical or virtual [15]. In a pure social setting, when no physical elements are required for hosting a human-robot interaction, the instantiation of the body is far less important than the social capabilities of the robot [16]. This leaves the question, what exactly is a social robot? In this section I will give a definition of a social robot and discuss the state-of-the-art of social robots in health care and in the field of child-robot interaction.

2.3.1. Definition of a social robot

A lot of definitions and taxonomies have been proposed trying to capture the concept of social robots [17]–[24]. The most commonly used definition of a social robot is given by Dautenhahn and Billard (1999):

> "*Social robots are embodied agents that are part of a heterogeneous group: a society of robots or humans. They are able to recognize each other and engage in social interactions, they possess histories (perceive and interpret the world in terms of their own experience), and they explicitly communicate with and learn from each other*" [17]

The term social interactive robots indicate the class of robots that socially interact with other social agents e.g. humans. The proposed taxonomies try to categorize social robots based on their ability to socially interact, perceive and interpret the world and their degree of embodiment (among other things). An often used taxonomy is the one proposed by Breazeal (2003) and extended by Fong et al. (2003). The robots in that taxonomy varies on one axis, the level of social interaction (7 levels) [18], [19]:

- *Socially evocative*: robots that passively encourage people to interact with it in a social way e.g. by relying on anthropomorphism.
- *Social interface*: robots that use human-like social cues and communication modalities to socially interact. Social behavior is processed at the interface level and therefore the social model of a person tends to be shallow.
- *socially receptive*: socially passive robots that benefit from the interaction e.g. by learning from it. A deeper social model is required than social interface robots.
- *Sociable*: socially participative robots that pro-actively engage in an interaction to satisfy internal goals and motivations. A deep social model is required.
- *Socially situated*: robots that are surrounded in a social environment which they can perceive and interact with. They can distinguish other social agents from non-social objects in the environment.
- *Socially embedded*: socially situated robots that interact with other social agents, are structurally coupled with their social environment and are (partially) aware of human interactional structures.
- *Socially intelligent*: socially embedded robots that possess deep models of human cognition and social competence.

The (my)PAL robot is at its current state a social interface. The research performed for the PAL project, including this thesis, is meant to work towards a sociable robot.

2.3.2. Socially Assistive Robots (ASR)

Due to the complexity of developing social robots most (attempts of) applications are introduced in a small domain and in a controlled setting. One of the most researched domains for social robotics is health care. It is important to note that most researchers and developers do not aim to replace the caregivers but provide them with a better toolset. These kinds of robots are often called socially assistive robots (ASR). The robots assist both the patient as the caregiver [23].

Socially assistive robots can have different functions and roles. They can for example be a motivator, educator, coach, companion or therapeutic play partner. For each of these roles there are different requirements regarding the social behavior [25].

Applications can be found in numerous health care domains targeted at different user groups. Most commonly targeted users are the elderly or children. ASR's can for example support elderly with mild cognitive impairments [26] or motivate healthy elderly to exercise more [27]. Furthermore, children with autism [28], [29] or disruptive behavior problems [30] can also benefit from receiving care from an ASR.

The (my)PAL robot can take different roles depending on the situation. The myPAL avatar is a motivator by having companion features. The PAL robot in the hospital is an educator by being a play partner.

2.3.3. Child-robot interaction (cHRI)

Anthropomorphism, attributing human characteristics to inanimate objects, is one of the driving psychological forces behind the successes of human-robot interaction. Children seem more eager to anthropomorphize robots than adults. This is strengthened by the tendency of children to play along, even if they know it is not as real as they pretend it to be. Furthermore, because children are still learning a lot about language, social protocols or even common sense they are either more oblivious or more forgiving towards the mistakes made by the robot. These properties make child-robot interaction fundamentally different from human-robot interaction with adults [31].

MyPAL will contain a cHRI user study where children will interact both with a physical as a virtual robot. Ros Espinoza et al. (2011), research performed for the ALIZ-E project, provide us with advice for doing a cHRI study [32]:

The first item to address are the legal and administrative issues and in particularly the rules, regulations and ethics regarding doing research with children. I will cover this in the next section. The second item is to be aware of impact the environmental setting of the experiment can have on the children. Children behave and respond differently towards the robot at home, at a camp or at the hospital. In the case of myPAL there are two environments to consider: the hospital and the home. This is further discussed in section 6.2.

The third item is the issue that working with children can be constraining for the experimenter. One issue is how to properly introduce the robot to the children because that influences how the children view the robot for the rest of the experiment tremendously. For example, if the robot falls during the introduction that participant might as well be ruled out of the experiment because it distorts the evaluation of the robot. Another thing regarding working with children is that briefing and debriefing needs to be carefully planned in order to manage the expectations of the children. Careful planning of the (de)briefing sessions also includes having a back-up plan when certain steps either go wrong or play out differently. The planning of the (de)briefing sessions of the myPAL experiment is discussed in detail in section 6.2.

The fourth advice is to plan the whole interaction meticulously. What does the robot (or avatar) say and when does he say it? The avatar behavior model is discussed in sections 4.4 and 5.1. The final item revolves around measuring and assessing HRI with children. Children must be able to understand the questions, e.g. in a questionnaire, without much effort, the question must be retrievable (taking into account the shorter memory span of children), questions must be as free as possible from a social desirable answer and the rating system in questionnaires must be easy to understand. Furthermore, a strong primacy effect is found among the children. They tend to copy their previous answer for answering the following question. The measures of the myPAL experiment are extensively discussed in section 6.2.

2.4. From doing research on children to doing research with children

Doing research that involves children requires a well-considered approach. Specifically, there are a number of ethical considerations that need to be addressed. How researchers view children, and

deal with these ethics, changed over time. In this chapter I will present a short overview of the historical changes and I will address the ethics related to research with children.

2.4.1. Historical overview

Childhood is a sociological concept rather than natural phenomena. How childhood is viewed changed a lot over time. Those view changes are closely linked to how the position of children in research changed. In the 15th century children were viewed as mini-adults who were born inherently evil. Raising them meant teaching them how to live a 'good' life. This changed at the beginning of the 17th when John Locke argued that children were born with a blank canvas that should be molded into something virtuous (nurture). At the end of the 17th century Jean-Jacques Rousseau argued that children were born inherently good and needed to be given the opportunity to express themselves freely (nature). This was followed by movement of more children's rights starting in the 18th century. Less work and more education followed for the children.

Children, or rather how they were viewed, evolved from frail and irrational beings that needed to be molded into recognized social actors operating actively and competently in their environment's. Furthermore, a more generational view was applied to the concept of childhood. Children were categorized based on their age and developmental phase. This segregation influenced the rights, deeds, economical participation, et cetera of the different groups of children. For example, the law stating how long children can work a day is a consequence of a generational view. Although these changes occurred worldwide there is still a great variability of how children are viewed and treated today within different cultures and nations.

In research children were first viewed as objects that could be studied. Research was done on children. It wasn't until after the second world war when this started to change. The atrocious studies done by the Nazi's involving children in concentration camps raised a lot of concerns. As the attitudes towards children started the change children were slowly more considered as participants. Research shifted from doing research on children to doing research with children [33].

2.4.2. Addressing the ethics

Why is it important to address the ethics regarding research with children explicitly? For one, children are one of the most vulnerable groups of participants. Ethics are our guiding principles that help us separate right from wrong, something we learn to do from childhood. While following a moral code might seem like common sense, the implications of research with children are not always as straightforward. The impact of research with children and how to deal with it is under constant debate. Not addressing the ethics properly can not only be harmful for the participants but also affect the trust bond between researchers and participants and ultimately the trust bond between science and society. Regardless of the size of a study one must address the ethics.

The first and foremost ethical considerations follow from these four core guiding principles: respect for autonomy, justice, beneficence and non-maleficence [33].

Children must consciously and actively decide that they want to participate. In order to safeguard this respect for autonomy informed consent is common practice. Discussing the experiment in advance and letting the participants sign a consent form provides a mechanism for children to become aware of their decision and giving them an opportunity to back out. Furthermore, allowing

them to back out at any moment in the experiment is another step to additionally respect their autonomy.

Justice can be categorized in three forms. The first is distributive justice, meaning that the available resources are equally distributed, e.g. all children get the same reward regardless of performance, gender, ethnical background, etc. The second and third are rights based justice and legal justice, meaning respect for children's rights and upholding the law. A core consequence is equality.

The research must have to be beneficial for either the children directly or society as a whole. Because participating with an experiment always comes with some form of risk. Besides the risk the children also invest time into the experiment. Therefore, it is important to consider the beneficence of the study and clearly document it. This thesis, but also the research proposal, is an example of that.

The last core principle is non-maleficence or, simply put, to do no harm. This might seem straightforward but should be considered thoughtfully nonetheless. Not only physical harm but psychological harm might be an unintended side-effect of the research. Take for example the introduction of a social robot in this research. The anthropomorphic appearance and the social capabilities of the robot triggers the children to form an attachment. When the experiment is over and the robot is taken away without taking this attachment into account it can harm the children. They might feel abandoned by the robot of feel they are losing a friend. Managing expectations and giving the opportunity to properly say goodbye reduces the changes for this to happen [34].

For doing human-robot interaction studies and specifically for doing child-robot interaction studies also ethical guidelines are being proposed [34], [35]. For example, Riek and Howard (2014) propose four ethical categories to consider: human dignity, design, legal and social. Robots must (be programmed to) respect the emotions, privacy, and frailty (both physical and psychological) of humans or in short respect the human dignity. Secondly robots must be designed as transparent, figuratively speaking, as possible. They must be predictable, show how reliable they are and have a kill-switch to turn them off. Besides following all the relevant rules and regulations all the robot behavior must be recordable and traceable for the purposes of litigation and dispute resolution. Finally, the social capabilities of the robot must be carefully developed. For example it must not be sexist or racist [35].

To safeguard whether my research respects all the ethical guidelines, general and applied to HRI, the research proposal was checked (and approved) by the internal review committee for experiments with participants of TNO and the medical-ethical committee of hospital Gelderse Vallei. The following standards are followed during the user studies:

- World Medical Association Declaration of Helsinki', revised during the 52th WMA General Assembly, Edinburgh, Scotland, October 2000 and the clarification on section 29 added to the WMA General Assembly, Washington, 2002.
- The quality guidelines of TNO (ISO 9001).
- Dutch laws regarding scientific medical research (WMO, 01-12-99).
- A Code of Ethics for the Human-Robot Interaction Profession (Riek and Howard, 2014, [35])

2.5. Situated Cognitive Engineering

Not all problems need to be solved using technology. Using technology as a solution just to use technology usually is not the best solutions at all. Carefully analyzing the problem and identifying possible solutions before looking at ways to implement the solutions, using technology or otherwise, is the best course of action. Situated Cognitive Engineering (sCE), developed by Mark Neerincx and colleagues, is a problems solving method that does just that [36]. In this section I will discuss the origin of sCE, how it works and the motivation of why it is used so centrally in this master thesis.

2.5.1. Origin of situated Cognitive Engineering

Situated Cognitive Engineering is derived from classical cognitive engineering. The term cognitive engineering surfaced in the 70's and 80's. Like many other fields it emerged from the zeitgeist of that period. Philosophers and scientist where in the process of verbalizing the fundamental differences between `the sciences of the artificial' and natural science [37]. More concretely, the number of developments in the field of computer science grew exponentially. First computers were used by a few scientists, now they were used by the general public at work or even at home. Scientists that were concerned with the understanding of interactions among humans and other elements of a system were trying to figure out how to incorporate this new system, i.e. a computer, into their models [38].

One of the founding fathers of cognitive engineering is cognitive scientist and usability engineer Donald A. Norman [39]. He is one of the first who articulated the cognitive engineering steps he took while dealing with modeling humans who were interacting with machines.

The basic concepts of cognitive engineering and its situated variant are the same. Namely, the subject under investigation is a cognitive system. This system consists of human and artificial agents that operate in the same domain and are delineated by roles, communications norms and protocols. The goal is to "develop systems that are easy to learn, are easy to use, and result in improved human-computer system performance" [39].

In classical cognitive engineering a cognitive system can be studied as an isolated object. In situated Cognitive Engineering the cognitive system is always viewed as embedded in a social, cultural and physical context. This adaptation has its origin in the rise of situated cognition [40].

The concept that objects should be viewed from within their context is as old as Aristotle. The famous silver chalice example by Heidegger illustrates how the four Aristotelian Causes can be applied to the question of how to view technology. The *material cause* relates to the material the chalice is made of. The *formal cause* defines the form the chalice has. The *efficient cause* relates to the what caused the chalice to become existent i.e. the silver smith. The *final cause* relates to the goal of the chalice [41]:

> *"Die causa finalis, der Zweck, z.B. der Opferdienst, durch den die benötigte Schale nach Form und Stoff bestimmt wird" – Martin Heidegger, Die Frage nach der Technik, 1954.*

In other words, the significance of a cognitive system is not only defined by itself (materials and form) but also by what its purpose is. In order to study a cognitive system properly `that what shapes its purpose', the context, needs to be included. A fundamental difference between classical

cognitive engineering and situated Cognitive Engineering is that sCE explicitly includes the context of the cognitive system. A second addition to the original method is the strengthening of relationship with technology [42].

2.5.2. The steps of situated Cognitive Engineering

Situated Cognitive Engineering consists out of five steps over which is constantly iterated to improve the result. The steps are: derive, specify, build, evaluate, refine (see Figure 2) [43].

2.5.2.1. Derive

The foundation of sCE is built from three components: operational goals, human factor knowledge and technology. These components need to be derived from the current objective. For example, if the current objective is to stimulate the self-management of a diabetic child, the derived main components are: motivate the children to keep a diary (operational goal) by satisfying the psychological needs that are required to be intrinsically motivated (human factor knowledge) using a responsive and adaptive avatar (technology).

2.5.2.2. Specify

The derived foundation is situated in a specific scenario. Using the operational goals and the scenario *use cases* are formulated. Use cases are to the point single goal descriptions of the behavior of the solution e.g. the avatar. For example, `the avatar responds emphatically to the current indication of the child's mood'. *Requirements* for the technology are derived from the use cases e.g. `the avatar is able to detect the mood of the child'. Whether this requirement is useful needs to be evaluated. This is done on the basis of a positive and a negative hypothesis which are called *claims* in sCE. A positive claim would be `children felt more connected to an empathic avatar (relatedness questionnaire) and used the diary more (time spend in the diary)'. The negative claim would be in the lines of children using the diary less when the avatar is present. Finally, *interaction design patterns* are ways to implement the requirements, e.g. the avatar will mirror the mood of the child as a form of empathy.

2.5.2.3. Build

The build step is straightforward: build the situated cognitive system. However, it is important to note that not only the evaluation step will result in useful information to refine the system but also the build step. Implicit assumptions may cause unexpected behaviors which leads, when discovered and explicated, to additional requirements.

2.5.2.4. Evaluate

Each requirement is evaluated based on the claims associated with it. This is called evaluation based on components. Claims are measurable statements. The claims formulated in the example above mention two metrics to validate them: a questionnaire to measure the perceived relatedness and the time spend using the diary. Furthermore, each operational goal is evaluated as well. This is called evaluation in the situated scenario, which is a more holistic evaluation. It could for example be evaluated whether all the combined responsive behaviors of the avatar lead to a more qualitative and quantitative use of the diary as a measure of the motivation of the child.

2.5.2.5. Refine

The last step is to use the outcome of the evaluation and build phases and make improvements throughout all the other steps. This is done multiple times making the process iterative.

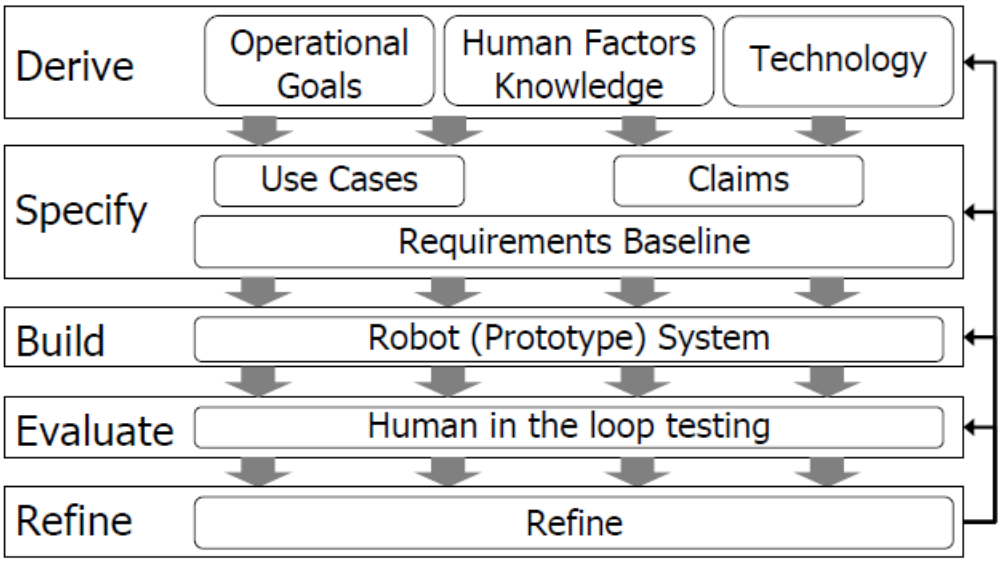


Figure 2: components of the situated Cognitive Engineering methodology. This image is based on Figure 1 from [43]

2.5.3. Why is situated Cognitive Engineering a central element?

The three foremost properties of sCE that lead me to choose it to structure the design, implementation and evaluation processes are the human-centered approach, iterative work style and broad scope.

Situated Cognitive Engineering is human-centered because the thorough problem analysis, followed by the identification of relevant human factors, lead up to introducing useful technological principles. Not the other way around. These three aspects form the foundation for solving the problem. Furthermore, the solution need not to have a complete technological nature. Solutions partly in the human domain are not only possible with sCE, they are usually the most effective ones. Take for example a generic overworked employee. A pure technological solution would be in the lines of increasing the employee efficiency with technological support. However, offering the employee a time-management course tackles the root of the problem. Without proper time-management the employee would just fill the freed time (by the technological solution) with other tasks. On the other hand, with all the time-management in the world one could not finish a job if the time to complete all the tasks exceeds the available time. The combination of increasing the employee's efficiency and time-management skills provide a more effective solution than one of them separately.

The second property, the iterative work style, is useful when dealing with a relatively novel problem or solution strategy. In both cased there are a lot of not well understood or even unknown facets. During the first design, implementation and evaluation cycle a lot of these unknown facets surface and new knowledge is generated. These experiences can greatly improve the solution found after that first cycle. Repeating that development cycle will incrementally work towards a better final solution. Instead of riding the waterfall once every step is repeated several times until the deadline is reached or the solution proves to be satisfactory.

A third property that stands out is the complete scope of situated Cognitive Engineering. It is not only a method to design a solution but it also provides structure for implementing the design and ultimately evaluating both design and implementation. In other words, sCE is a method that supports the complete developmental process. In practice the results found in the evaluation link not only to the implementation but also to the design and its foundation. This helps to improve the overall project tremendously. This complete approach is a requirement for an effective iterative work style [44]. All three properties are especially useful for managing and improving complex projects with novel technology such as the one defined in this master thesis.

3. Foundation

The aim for this project, as introduced in chapter 1, is to develop a digital diabetes diary where the avatar autonomously employs multiple behavioral strategies to motivate the children for a longer period of time. In this chapter the operational demands for such a system are extracted from a concrete problem scenario. The operational demands help to identify which human factors and technological knowledge are relevant for solving the problem. Together these elements form the foundation of the solution. The foundation will lead to two types of questions: design questions and research question. The former revolve around how to design the desired system given the knowledge present in the foundation and the latter are about investigating how children interact with the system for longer period of time. Answering these questions will strengthen the foundation for a future iteration of the design cycle.

3.1. Operational demands

Scenarios are a key component within situated Cognitive Engineering. Not only do they provide a concrete example of the problem they also take the perspective of all important stakeholders. This helps to find a solution that is beneficial for all parties. In this section a scenario is provided where the lack of motivation is the main problem. From the scenario the stakeholders and demands for possible solutions are identified.

3.1.1. Problem scenario

Frank (age 9) was diagnosed with diabetes three years ago. Frank and his parents are at the hospital. Today they meet with various medical professionals, amongst others the doctor, the diabetes nurse, and a dietician. Together they try to paint a picture of the current state of Frank's diabetes. From there the medical professionals can adjust the treatment plan and give Frank recommendations to improve his well-being. This time he receives the recommendation to keep a diabetes diary. Frank is told that not only the medical professionals would gain more insight into the developments of his diabetes but also Frank himself. The diabetes diary covers:

- blood glucose levels;
- insulin: regular and additional;
- outcome from bolus calculator;
- amount of carbohydrate in meals;
- classical diary elements:
    - activities such as school, sports, playing with friends and
    - feelings (emotions) during those activities.

Some of these values are uploaded automatically to the diary by the measuring equipment but others need to be entered manually.

Back at home Frank does not often use his diary. Frank's parents try to push him to record his values, which only causes Frank to not want to do it even more. Eventually the parents update the diary themselves a couple of days before the next appointment at the hospital. Frank does not really care for keeping a diary because he does not see the importance of it. He furthermore considers it an annoying task to do. He often forgets to look at his diary entirely let alone update it. In other words, Frank is not motivated to keep a diary. This makes it harder for Frank to gain insight in his diabetes (and ultimately manage it better) and for his parents and medical

professionals to monitor Frank's condition. The medical professionals do not have sufficient (reliable) information about the developments of Frank's diabetes, wellbeing and self-management skills (over longer periods of time and in specific moments) [45].

3.1.2. Problem analysis

From the scenario the stakeholders, their values, and other possible actors (or secondary stakeholders) can be identified. Furthermore, the scenario can be broken down in core components.

Stakeholders

'Frank' represents children between 7 and 12 years old who are diagnosed with Diabetes Mellitus type I. They have learned the basic skills to manage their diabetes and would benefit from keeping a digital diabetes diary. It can help children to identify their diabetes related experiences, e.g. dizziness, lack of concentration and mood swings, understand the cause, e.g. low glucose level, and gain insight *in* how to manage it, e.g. consume glucose tablets.
Values: feeling well, autonomy and having fun.

'Frank's parents' represents the parents of diabetic children. They have learned the basic skills to help their children manage their diabetes. The diary can help them to better understand what their child is going through.
Values: healthy child and being supporting/helpful

The medical professionals can use the information collected with the diary as a starting point for a discussion with the children about their experiences and to adjust the treatment plan.
Values: healthy child, caregiving and having all the information

Other possible actors

Other possible actors are family members, friends of the children and friends of the parents. The parents can possibly use the information from the diary to explain the condition of their child to family members and friends. The children themselves can use the diary as well to show their friends and class mates what diabetes is all about.

Breakdown

The problem scenario can be broken down into concrete appointable problems, their causes, other relevant information and goals to overcome the identified problems.

Problems

- The children do not (want to) consult and update the diary.
- The children forget to consult and update the diary.
- The children will not gain insight in their diabetes as fast as possible (as with the use of the diary would have been the case).
- The caregivers (parents and medical professionals) are inhibited in monitoring the condition of the child.

Causes

- The children are not motivated to keep a diary.
- The children do not see the importance of the diary.

- The children consider keeping the diary an annoying and time-consuming task.
- The parents try to help the children too much (taking away their responsibility).

Other information

- The children are coerced to update the diary (with counterproductive consequences).
- The parents fill in parts of the diary instead of the children.

Demands

- Motivate the children to use the diary regularly.
- Let the children understand the importance of the diary.
- Make consulting and updating the diary more fun.
- Help the children to gain insight in their diabetes using the diary.
- Help the caregivers to monitor their child using the diary.
- Enable parents to help their child without having to coerce.

3.2. Human factors

When children can be moved to keep a diary they are considered to be motivated. People can be moved because it is inherently enjoyable or because it leads to an appointable outcome. These two forms are called intrinsic and extrinsic motivation respectively [46]. The most straightforward approach, when having an avatar present in the diary, is to let it explicitly encourage the child. However, this form of stimulating the motivation of children might not be the most effective one. In this section I will introduce the Self-Determination Theory (SDT), a framework for intrinsic motivation, and use it to derive relevant human factors that can be used to meet the operational demands.

The Self-Determination Theory is a meta-theory that encapsulates several theories about intrinsic and extrinsic motivational phenomena found in numerus field and lab studies. SDT functions as a broad framework that allows researchers to position motivational studies, provides a formal theory that defines sources of motivation, and describes the role of intrinsic and extrinsic motivation within cognitive and social development [47].

SDT was initially developed by Edward Deci and Richard Ryan in the end of the 70's and beginning of the 80's. It was not until the mid-80's the theory was more generally accepted and lead to an explosive increase in studies around it by other researchers. Applications, and research, of the theory are found in the areas of education, healthcare, relationships, goals, sport and exercise, psychotherapy, et cetera [48].

The Self-Determination theory focusses on how people's sense of initiative is stimulated or inhibited by cognitive, social and cultural factors. SDT argues that when someone's individual experience of autonomy, competence, and relatedness is supported it positively influences the intrinsic motivation of that individual to engage in the activities at hand. Executing an activity under the supporting condition result in an enhancement in performance, persistence and creativity [49]. Most of the available research revolves around adults but the self-determination theory is also validated for children [50].

The question I will explore now is which human factors, given the scenario of keeping a diary, can play a role in supporting the need for autonomy, competence, and relatedness?

3.2.1. Autonomy support

Let's start with autonomy. Research shows that supporting autonomy in contrast to external motivation e.g. using a reward or punishment is more effective and makes children happier [51]. Furthermore, The Good Childhood Report, a large survey study, confirmed the importance for autonomy in the form of freedom of choice on the well-being of a child [52].

However, the amount and the form of the freedom children are given in their choices affects the effectiveness of the autonomy support. A balance needs to be found between total freedom and providing structure in the decision making. Setting a relevant goal is a useful concept for controlling that balance [53]. Not only the format of how a goal should look like but also how it is evaluated is important for autonomy support. Being evaluated in a competitive manner or by strict rules enforced by an authority figure is far less beneficial for supporting autonomy than co-operative goals. Finally goals should only be assessed ipsatively, i.e. based on one's own results, and not based on the results of others or a golden standard [54].

A different framework for autonomy support is presented by Stefanou et al. (2004). When children are, for example, presented with a task in the classroom and they are allowed to make decisions about what to do exactly their *organizational autonomy* is supported. When the children are given ownership over the process of completing a task their *procedural autonomy* is supported. When children are afforded the opportunity to evaluate the outcome ipsatively their *cognitive autonomy* is supported [55]. To translate this to scenario at hand the children need to be given ownership over the added content (organizational autonomy), the way they add the content and when they add the content (both procedural autonomy).

The identified human factors regarding autonomy support brings about the following requirements:

AUT1. myPAL lets the children set their own goals that only relate to themselves.
AUT2. myPAL balances freedom and the structure within the goal setting process appropriately.
AUT3. myPAL encourages, but not forces, the child to achieve goals from a co-operative perspective e.g. by cheering him/her on
AUT4. myPAL evaluates the goals ipsatively.
AUT5. myPAL applies appropriate balance between freedom and structure to the process of adding content.
AUT6. myPAL gives the children ownership over the added content and how and when it is added.

3.2.2. Competence support

A way to increase the feeling of competence is to praise the children when they show good or wanted behavior [56]. The preferred behavior is the amount of content and the consistency of adding it. The higher the better in both cases. MyPAL should praise children who adds enough content and do that consistently. For example, when they use the diary for more than one day in a row.

Furthermore, the desired behavior is a valuable source to structure the goals. Goals should focus on adding content consistently. More importantly, the children who obtain those goals should be

praised on their success as well. Finally, viewing all your achieved goals has also a positive impact on you feeling of competence [57].

The following requirements can be extracted from these human factors:

COM1. myPAL praises children who add content and who do that consistently (including setting goals).
COM2. myPAL praises children who reached a goal.
COM3. myPAL shows all the reached goals to the child.

3.2.3. Relatedness support

Supporting relatedness between myPAL and the child requires a direct interaction with the embedded avatar. Bonding with a virtual robot has been found to increase the motivation to keep a digital diabetes diary [5]. The social capabilities of a (virtual) robot is an important factor for feeling related to it. The myPAL avatar will be perceived as a social entity when it displays forms of empathy, solidarity, humor and curiosity. These aspects can all be implemented with the avatar using a combination of gestures and speech as done in [16].

After the 'why', 'what' and 'how' questions the 'when' question follows. When would it be appropriate to engage in a conversation. An avatar that replies to the added content was found to be much appreciated by the children [5]. It gives the child the feeling that someone is listening to them. Besides the added content itself provides the necessary input to generate a relevant conversation starter. When myPAL would collect the mood of the child it could be matched by the avatar during its reply. Matching the mood of the child is a concrete way to show empathy [58].

A different tactic that even has a more direct effect on how much and how consistent content is added is the phenomena of mutual self-disclosure. When the avatar shares information and content, e.g. pictures, about itself the children are also more inclined to do so. Furthermore, they feel more related to an avatar that shares about itself [5].

The identified human factors that support building a prosperous relationship with the avatar result in the following requirements:

REL1. myPAL shall reply socially to the added content by showing signs of empathy, solidarity, humor and curiosity.
REL2. myPAL shall match the mood of the child during its reply.
REL3. myPAL shall disclose information and pictures about itself.

3.3.Technological principles

In myPAL there are three relevant technological principles involved. The first is an online digital diabetes diary. The second is the avatar that is embedded in the diary. The avatar is based on an actual robot that also has a role in the project. The global properties of the technologies are discussed as well as their expected benefits and downsides.

3.3.1. Online digital diabetes diary

To provide children with a convenient way of supplying the specified information an online digital diabetes diary will be used. It is a web based graphical platform containing at least a page for

- logging in;

- creating, reading, updating and deleting (CRUD) activities, pictures, measurements and goals and
- providing access to all the added content and interactional data for a registered researcher.

Furthermore, the platform hosts a virtual avatar that will be present on (almost) all the user pages.

Expected benefits

- Can be accessed virtually everywhere with a computer, tablet or smartphone so long as there is an Internet connection.
- There are lots of existing digital diabetes diaries available for inspiration and best practices.

Expected downsides

- Each glucose measuring device manufacturer has its own diary/logging system. It will be a challenge to be able to collect all the required data automatically from all different devices.
- There does not exists a diary that meets all the human-factor requirements. A custom-made diary must be developed.

3.3.2. Avatar

Because it is impractical and expensive to give every child a robot, the children will only meet the robot at the hospital or during group activities, e.g. a diabetes camp. In order to still use the beneficial aspects of a social humanoid robot and the constructed relationship between the child and the robot, a virtual representation of the robot will be used. The avatar of the robot will be present in the various heath apps of the PAL-project, of which the diary is one. Research shows that when no physical aspects are required a virtual robot could elicit the same responds in people. The sociability of the (virtual) robot is in that case a more important predictor of its effectiveness and likability [16].

The virtual avatar can speak and display text and images on the screen. It does not have speech recognition. It processes the incoming data from the diary entries, e.g. activity descriptions, indications of mood and the information that is provided with the uploaded pictures, and subsequently selects a response (= gesture + speech) that contributes to either the feeling of competence, autonomy, relatedness or a combination.

Expected benefits

- Easier controllable than a physical robot (less psychical constraints).

Expected downsides

- It is unclear whether the relationship between the child and robot is actually transferrable to the virtual avatar.

3.3.3. Robot

A social humanoid robot has beneficial effects on the self-management of children with diabetes type I. A robot is able to motivate children to undergo useful activities, e.g. play a diabetes quiz (educate) or keep a diary (collect and reflect on diabetes related values and moods), and make these activities more fun. Ultimately, this supports their insight in diabetes [5], [59], [60]. A social humanoid robot is a central part of the PAL-project. It however will play a minor role in the myPAL project. During the user study (see chapter 6) children will meet the physical robot, that serves as

the model for the avatar, so that they can link the avatar and robot together. Whether children actually view the robot and the avatar as one entity is left for future work. The physical robot is introduced to the children in the myPAL evaluation in order to match the experimental scenario to the intended use of the complete PAL system.

Expected benefits

- Social humanoid robots are more effective to fulfill the needs for competence, autonomy and relatedness than a disembodied computer program.
- There are suitable commercial robot platforms available that are easy to use.

Expected downsides

- Social humanoid robots elicit (too) high expectations from its users that need to be managed in order to prevent disappointments (and a drop in effectiveness).
- Designing and maintaining a human-robot interaction can be a cumbersome task.
- Effects of gender, age, personality, general attitude towards robots, novelty and culture have been reported to influence the way robots are perceived by its users [16].

3.4. Design and research questions

The purpose of my master thesis project is twofold. The first is to develop a digital diabetes diary with an autonomous avatar that is able to motivate children with diabetes to use that diary structurally. The second is to research how children with diabetes behave towards myPAL for a certain period of time. This means that this thesis describes two interchanging research lines. The first is the applied research line containing design questions about how to incorporate the requirements and knowledge identified by the foundation into an actual functioning prototype. The second is the fundamental research line where the behaviors and motivations of diabetic children that interact with the myPAL are investigated.

3.4.1. Design questions

The development process is structured by the situated Cognitive Engineering method. As a consequence, the design questions are as well shaped by this method. Given the foundation defined in chapter 0 and the structure situated Cognitive Engineering provides, the design questions are:

DQ1. What are the core functions of a digital diabetes diary with an autonomous avatar that is able to motivate children with diabetes to use that diary structurally?

DQ2. What avatar behaviors and dialogue structures contribute to increasing the (intrinsic) motivation of children with diabetes to use the diabetes diary structurally?

Chapters 4 and 5 discuss the functional and technical design considerations of myPAL focusing on answering both questions. Note that this thesis is not meant to be a complete functional and technical design for the whole software. Although a complete overview is given only the relevant parts, for answering the design questions, is discusses in full detail.

3.4.2. Research questions

If the previous research, described in the foundation, shows us anything it is that people interact differently with robots, avatars or software systems. Although certain properties show good results in general, e.g. a self-disclosing agent invokes more self-disclosure from the children, this

need not be the case for every child in particular nor need it to be consistent over a longer period of time. Studies investigating reoccurring interactions over a period of weeks between children and an autonomous embedded avatar are rare. With this thesis I aim to contribute to this novel line of research. A step towards knowing more about long-term child-robot interactions is performing an experiment that extends the usual one-time interaction. To evaluate this longer term experiment I aim to answer the following research questions:

RQ1. Which avatar behaviors are observed by the children?
RQ2. What is the effect of the amount of observed behavior on attitude towards the avatar?
RQ3. What is the effect of attitude towards the avatar on the antecedents of motivation, motivation itself and performance?
RQ4. What is the effect of the antecedents of motivation on motivation and performance?
RQ5. What is the effect of motivation on performance?

And a step towards a personalized system is identifying categories of children that interact differently from each other but show similarities within the category. This step is paired with the following research questions:

RQ6. What is the effect of user demographics on the amount of observed behavior, user attitude, antecedents of motivation, motivation, and performance?
RQ7. What categories of children can be identified based on the performance?
RQ8. What categories of children can be identified based on the development of motivation?
RQ9. What are the characteristics of the identified categories in terms of performance, motivation, demographics, and attitude towards the avatar?

During the user study various types of behavioral data about the use of myPAL and the added content is collected to identify different patterns of interaction. Furthermore, the motivation to use the diary is measured at the start, halfway and at the end of the study to monitor the development of the motivation. Lastly, the attitudes towards the avatar are measured during the evaluation session.

This research has an explorative nature due to the novelty of the used methods and the small population of possible participants. As a consequence:

- all the participants interact with the diary and avatar under the same conditions;
- instead of focusing on a few features of the system the broad range of features are all tested as a whole and
- the statistical power to find categories of similar interacting children that differ significantly might be too low.

Therefore, the results of the user study are used to form design recommendations and hypotheses for future work.

4. System Design Specifications

A broad foundation is constructed in the previous chapters that provides us with requirements for the system. In this chapter the functional specifications of myPAL are defined and discussed in order to build a prototype (see the next chapter). The system design specifications answer the two design related questions posed in section 3.4.1 by making important design decisions explicit.

This chapter is structured as follows. First the desired operations of the system are described in the form of design scenario's. From these scenario's use cases are derived. This is followed by a decomposition of the use cases into functions, and more specifically, interaction design patterns. Each function has a desired effect on the user. Finally, measures to detect these effects are discussed in the last section.

4.1. Design scenario's

The myPAL system should be used, eventually, by all the relevant stakeholders such as the child, their parents, the medical professional and maybe a system administrator. To keep this first iteration of the myPAL system feasible only two users are facilitated by the system. Those users are the primary user, the children, and the researchers / developers. From the perspective of both users a design scenario is developed.

4.1.1. "PAL stimulates the motivation of Frank to keep his diary"

Before Frank goes to bed he is eager to log in to myPAL, an online platform that hosts Frank's digital diabetes diary and the virtual avatar of *PAL*. PAL is the social humanoid robot Frank has met during his stay in the hospital. Frank logs in to myPAL every evening.

Frank uploads and/or manually adds the daily values of his glucose measurements and insulin intake to the diary. Furthermore, using the diary Frank constructs a timeline of his day. He does that by adding activities, e.g. school, sport or meal, to the timeline. Frank adds pictures he has taken that day to some of those activities. Each activity requires at least the following information: type of activity (e.g. school, sport, meal, or something else), an activity title, the date and time, a short description and an indication of the mood (happy, neutral, or sad) at the time of the activity. Adding a pictures is optional.

When frank adds an activity the avatar almost always responds to the content of the activity. The avatar responds happy to an activity where Frank had a good mood and the avatar responds supportive to an activity where Frank had a bad mood. A response consists of a bodily gesture and a verbal utterance, e.g. both arms go up in the air while saying "You did it, well done!".

The avatar shares photos of the PAL robot undergoing various activities fitting with the persona of the PAL robot. Furthermore, Frank and the avatar regularly set goals together that Frank tries to complete, e.g. add an activity to the diary for three consecutive days. When Frank completes a goal the avatar complements Frank.

Frank likes the responses of the avatar to his values, activities and photos. The main reason why he likes to log in to the diary is the presence of the avatar. Frank feels more competent doing his measurements and sharing them in the diary. Furthermore, he likes he can set his own goals together with the avatar. All in all, Frank is more motivated to keep a diary than before.

4.1.2. "PAL provides a researcher easy access to the user added content and control over the avatar"

When a researcher logs in to the PAL system he or she will be able to view, add, update or delete users. There is an overview page that displays all the current users and provides a form to add new users. Furthermore, myPAL provides easy access to all the added content (activities, measurements, goals and pictures) per user as well as all the interactional data (user behavior e.g. which buttons were clicked and avatar behavior e.g. how did the avatar responded to an added activity). When the researcher clicks on a user, on the user overview page, a data view page for the selected user is displayed. All the mentioned data is displayed on this page.

The researcher will also be able to view, add or delete avatar behaviors and gestures. Separate pages are available for these functionalities. A behavior consists of verbal element and a gesture. When adding a new behavior, the researcher enters the main text and all the alternatives for the verbal part and selects an existing gesture. A new gesture, a video file, can be uploaded on a separate page.

4.2. Use cases

From both design scenario's use cases are derived that together will cover all the functionalities of the myPAL system. The use cases that displayed in Figure 3 are derived from the first scenario. The use cases are divided in five categories: general system cases (blue) and cases related to activities (green), pictures (yellow), measurements (red) and goals (orange).

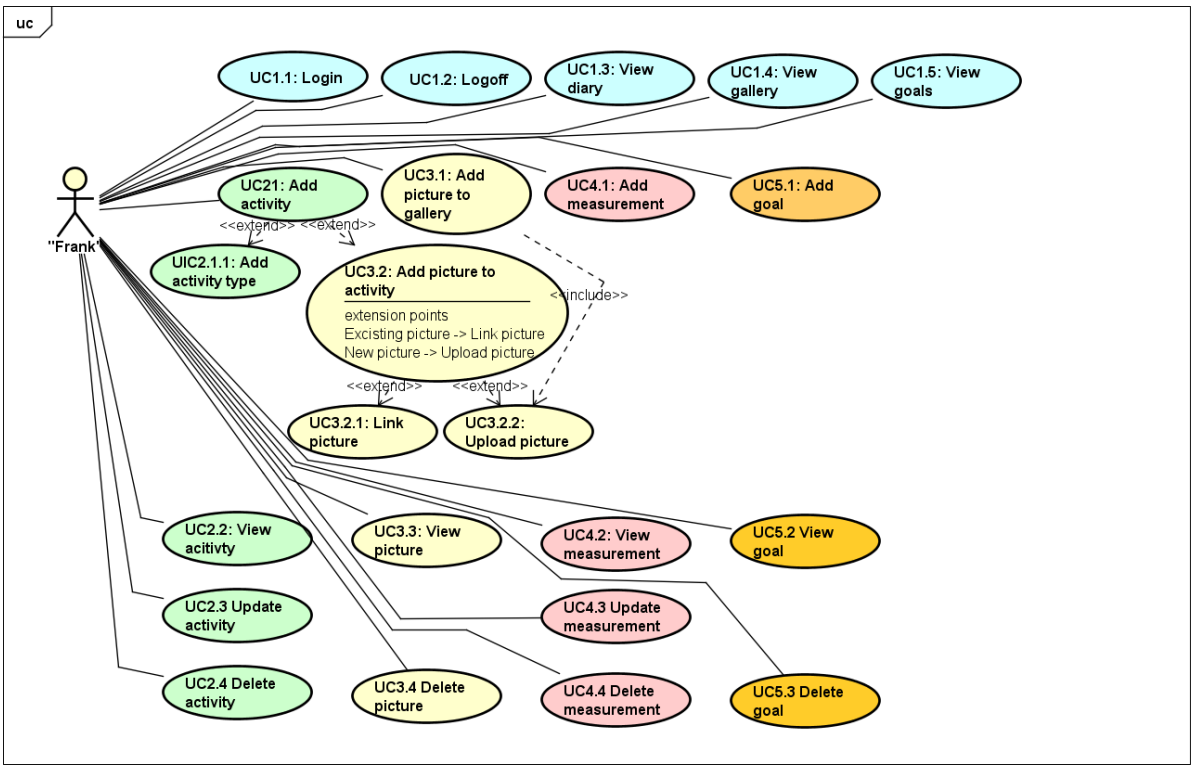


Figure 3: use case diagram representing the "PAL stimulates the motivation of Frank to keep his diary" scenario. There are five types of functionalities in the system indicated by the different colors. The blue use cases describe the general system functionalities, the activity related use cases are green, the picture related use cases are yellow, the measurement related use cases are red and the orange indicated use cases are related to goals.

The larger part of the use cases revolves around CRUD operations and follow the standard pathway. The use case of a standard CRUD operation on the web contains the following elements: accessing the relevant page, perform the operation of creating, reading, updating or deleting, receiving feedback about the success or failure of the operation and a redirect to a follow-up page. An addition to the use cases that perform deleting operations is that the system (not the avatar) prompts the user with a questions asking confirmation (e.g. are you sure you want to delete this?). In case of myPAL each CRUD operation is logged. To prevent redundant information only the use cases that contain an interaction between the child and the avatar are described in a use case description below.

Use Case	**UC-F 1.1 Logging in**
Summary	The child logs in by entering his/her credentials and is greeted by the myPAL avatar.
Actor	Child
Pre-condition	Child is not logged in; login page is loaded.
Main scenario	1. The child enters his/her name and password in the input field and presses the button login. 2. The credentials are correct and the diary main page is loaded. 3. The logging in user action is logged. 4. The avatar greets the child e.g. by saying "Hi, nice to see you again". 5. The greeting avatar action is logged.
Post-condition	The child is logged in, the diary main page is logged in and both the user and avatar actions are logged.
Alternative	2a. 1. The credentials are not correct. 2a. 2. A message is displayed on the login page stating that the credentials are not correct and that a new attempt should be tried. 2a. 3. Go back to main scenario 1.

Use Case	**UC-F 2.1 Add activity**
Summary	The child can follow two routes to add an activity: with or without the avatar. When the child has provided all the required information the activity is saved in the diary.
Actor	Child
Pre-condition	The child is logged in; the main diary page is loaded.
Main scenario	1. The child presses the add activity button. This user action is logged by myPAL. 2. The avatar asks the child whether he/she wants to add the activity together with the avatar or by himself/herself. The avatar behavior and the user decision is logged by myPAL. 3. The child is prompted to select an activity type, supply a title, date, time, description and select a mood (happy, neutral, sad) in the form of a

	smiley. Uploading a new picture or linking an existing (unlinked) picture is presented as optional. a. When the child chooses to add the activity together each item is queried separately. The avatar asks the child to supply the item in the displayed input field. b. When the child chooses to add the activity alone all the items are presented in a single form that needs to be filled in. 4. The child entered all the information and presses the submit button. 5. All the information meets all the constraints and myPAL generates a new activity. This user action is logged by myPAL. 6. The diary page, set on the date of the added activity, is displayed. 7. Based on the avatar behavior model (taking the activity type and the mood as input) the avatar responds to the added activity. This avatar behavior is logged by myPAL. a. The avatar can make a statement. b. The avatar can ask a question, provide options to answer the question and reply to the selected answer. c. The avatar can ask an open question and reply in general. d. The avatar can share information about himself by showing a picture with context.
Post-condition	A new activity is created and saved; the diary page, set on the date of the added activity, is displayed and all the user and avatar actions are logged.
Alternative	4a. 1. One or more information items do not meet the requirements. 4a. 2. The single form page (as in 3b) is displayed showing the error messages below the items in question. 4a. 3a. The child corrects the mistakes. Return to main scenario 4. 4a. 3b. The child cancels the adding process of the activity. This use case is terminated unsuccessfully and the main diary page is loaded.

Use Case	**UC-F 1.5 View goals**
Summary	When the user accesses the goal page the avatar will encourage the child either to add a new goal (when none are active) or complete an active goal
Actor	Child
Pre-condition	Child is logged in
Main scenario	1. Child access the goal page. 2. The goal page is displayed and the user action is logged. 3. a) When there are no goals currently active the avatar encourages the child to add a new one. It randomly chooses to recommend a daily or a total goal. b) When there are active goals the avatar encourages the child to complete one of them. The avatar randomly selects a goal to give encouragements about.

	4. The avatar action is logged.
Post-condition	The goal page is displayed and both the user and avatar actions are logged.
Alternative	

Use Case	UC-F 5.1 Add goal
Summary	User either adds a daily goal or a total goal to myPAL
Actor	Child
Pre-condition	Child is logged in. The goal page is displayed.
Main scenario	1. The child presses either the 'add daily goal' or 'add total goal' button. 2. The appropriate add goal page is loaded and displayed. 3. The child selects what goal type he/she wants to add. a. Daily goal: add x activity / measurements / pictures / activities from yesterday. b. Total goal: log in for y consecutive days / add at least one activity / measurement / picture for y consecutive days. 4. The child selects the goal parameter (integer x or y). a. Daily goal: x ranges from 1 to 10. b. Total goal: y ranges from 1 to 7. 5. The child presses the add goal button. 6. The goal is saved by myPAL and the user action is logged. 7. The goal page is displayed now showing the newly added goal. 8. The avatar praises the child for adding a goal. 9. MyPAL logs the avatar action.
Post-condition	A goal is added and displayed on the goal page, the child is praised by the avatar and the user and avatar actions are logged.
Alternative	5a. 1. The child presses the cancel button. 5a. 2. The use case is terminated and the goal page is displayed with no changes.

In Figure 4 the use cases extracted from the second scenario, representing the researchers' perspective on myPAL, are displayed. The general use cases, logging in and off and viewing all the main pages, are marked in blue. The other cases are about managing the users (green), avatar behaviors (yellow) and avatar gestures (red). The researcher does not interact with the avatar. The structure of the blue use cases is as follows: user presses a button (e.g. logoff, user page, etc.) and myPAL serves the appropriate response (e.g. logging off, displaying the user page, etc.). The management use cases (green, yellow and red) follow the standard CRUD pathways as described above.

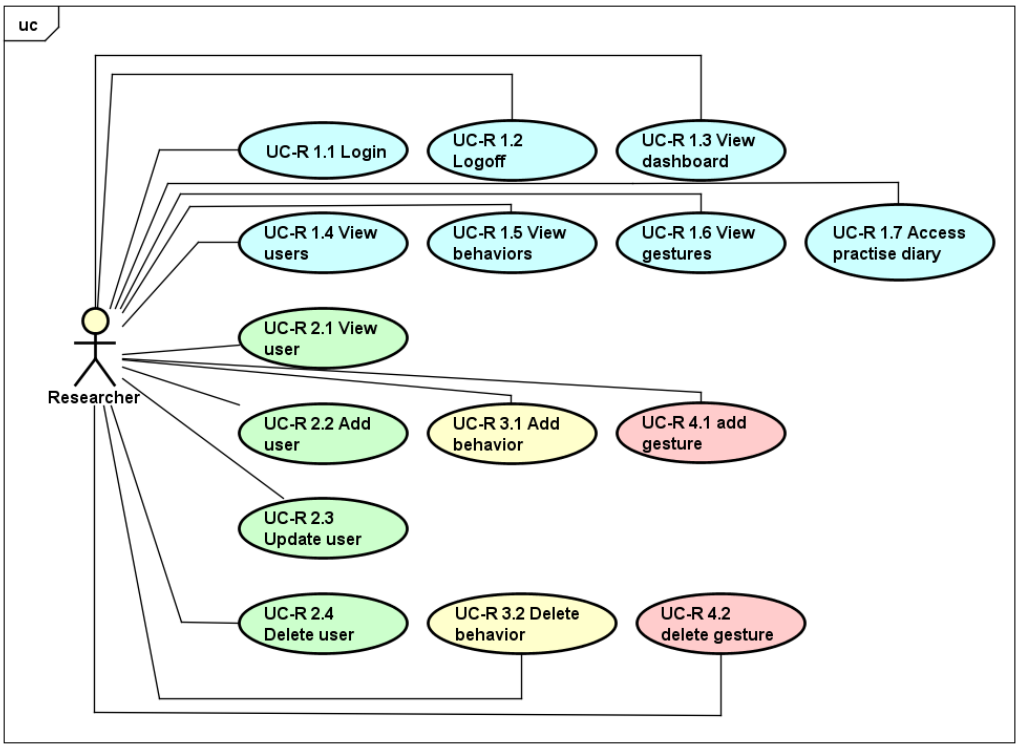


Figure 4: Use case diagram representing the "PAL provides a researcher easy access to the user added content and control over the avatar"-scenario. The blue cases represent the main functionalities, the green cases are about user management, the yellow and red are about respectively managing the avatar behaviors and gestures.

4.3. Functions

The focus of this thesis does not lie on the software engineering aspect of myPAL but on the design and research questions formulated in section 3.4. This section is not meant to be a complete functional design. Its purpose is to answer design questions DQ1 and find all the core functions related to stimulating the motivation the children to keep the diary on a regular basis. Furthermore, in section 3.2. human factor requirements that these functions should fulfill are identified. All the functions presented here should cover all the requirements together.

Table 1 contains all the relevant functions and their relation to the interaction design patterns, which are discussed in more detail in section 4.4, and the human factors requirements they fulfill. The first set of functions that are required cover the basic application functions, such as displaying a page, the logging of user and avatar actions, for research purposes, and the create, read, update and delete (CRUD) functions for managing the content. When we talk about keeping the digital diabetes diary we mean creating and reading activities, measurements and pictures. Updating or deleting these items is important for three reasons.

The first two revolve around ownership of the content. I consider it a fundamental right that the content the children add remains theirs and that we should treat that ownership with respect. Although this is a research project and you want to know exactly what the children add, we also have the responsibility to protect the privacy of the children. Therefore, when something is altered or removed in myPAL it really is altered or removed without keeping an original copy. In myPAL the privacy and ownership of the children is protected by design as defined by [61].

Having ownership over the items you add and the way you add them furthermore supports the feeling of autonomy as stated by requirement AUT6. The third reason is an argument with a more practical nature. Updating and deleting functionalities are necessary when children want to correct a mistake in their added content. It improves the overall usability of the system.

A different form of autonomy support is provided when children are given ownership over how they want to add the content (procedural autonomy support – AUT6). myPAL provides two paths, one with the avatar and one without. The avatar asks the children every time they want to add something if they want to do it together or by themselves.

To provide balance between structure and freedom for adding content (AUT5) a format is provided for children that they have to fill in. For example, for adding an activity children need to select an activity type from an existing list (or add a new one – AUT6), enter a title, description, etc. What information they need to supply is fixed, but how and how much they supply is up to them. The only optional activity item is adding a picture.

The main source for relatedness support is put in the responses the avatar gives after an activity is added. Note that the avatar does not respond after a measurement is added. The avatar matches to mood of the child (AUT2), shows signs of praise (COM1), empathy, solidarity, humor and curiosity (REL1) in the reply and shares information, mostly in the form of pictures, about itself (REL3).

In the goal section of myPAL the avatar cheers the children on to either add a new goal or complete an existing goal (AUT3). Creating a goal is on its own a way for children to express what they themselves think they achieve which should give a positive boost to their autonomy (AUT1). There are two types of goals that the children can choose from, daily goals and total goals. Within both types of goals, the children can choose different targets for those goals, e.g. directed at adding activities or pictures. The difficulty of the goal the last element the children can select. In order words, myPAL provides the necessary structure while still giving enough freedom of choice (AUT2).

After adding a new goal, the avatar praises the children (COM1) as well as after a successful goal (COM2). A goal is deemed successful when the set target is reached before the set deadline. These are components the children control themselves. The goals are not compared with a golden standard or other participants (AUT4). Finally, all the met goals are displayed on the goal page for the children to see (COM3).

Functions	Purpose	Interaction design pattern	Human Factor Requirement
Display page, process button press, authenticate user	Basic application functionalities	-	-
Log user and avatar actions	Research	-	-
Create & read activity, measurement and picture	"Keeping the diary"	-	-
Update & delete activity, measurement and picture	Provide organizational ownership	-	AUT6
Ask to add activity together or alone	Provide procedural ownership	Question with answer options	AUT6

Add activity together	Provide structure with guidance	Ask activity item	AUT5
Add activity alone	Provide structure without guidance	-	AUT5
Create, read, update and delete activity type	Provide organizational ownership	-	AUT6
Respond to added activity	Relatedness support	- Mood matching - Question with answer options - Open question - Remark - Self-disclosure	REL2 REL1 REL1 REL1 REL3
	Competence support	- Praise	COM1
Encourage adding a goal	Stimulate external motivation	Encourage	AUT3
Encourage completing active goal	Stimulate external motivation	Encourage	AUT3
Create goal	Autonomy support	-	AUT1, AUT2
Praise adding a goal	Competence support	Praise	COM1
Praise completing a goal	Competence support	Praise	COM2, AUT4
Read goal	Competence support	-	COM3
Delete goal	Provide organizational ownership	-	AUT6
Legend	Avatar	Web application	

Table 1: Functional table of myPAL that lists all the relevant functions and their relation to the interaction design patterns and human factor requirements. The blue rows are functions executed by the avatar and the white rows are executed by the web application.

4.4. Interaction Design Patterns

Some of the functions identified in the previous section include reoccurring interactions between the child and the avatar. These interaction design patterns identified in Table 1 are discussed in more detail here. Although technically speaking situated Cognitive Engineering allows for a broader interpretation of interaction, e.g. between the web application and the user, only the avatar related interaction are discussed here in order to focus on design question DQ2. DQ2 asks the following question: "What avatar behaviors and dialogue structures contribute to increasing the (intrinsic) motivation of children with diabetes to use the diabetes diary structurally?"

Avatar behaviors are composed of three components: speech (audiovisual), gesture and content. An audio file is played to the user and the text is displayed on the screen. Gestures are video clips of the avatar moving, e.g. performing a waving motion. The content component is a collection of different things. When the speech component is a question the content component provides ways

to answer that question by either presenting buttons with the answer options on it or a text input field to type the answer in. The content component can furthermore consist of something the avatar wishes to share such as a picture. Finally, the content component can also be empty.

The interaction design patterns listed in table 1 are discussed in more detail below.

4.4.1. Remark

A remark is the most basic pattern and has the function to relay a simple text to the child. The text usually consists of one sentence. Most remarks are meant to express interest in the child, empathy, solidarity, curiosity or be funny. The text, in audio visual form, is accompanied by a fitting gesture. Fitting with the text and the mood (see mood matching). There is no additional content element displayed with this pattern.

4.4.2. Question with answer options

The 'question with answer options' pattern is similar to the remark pattern only now the text is formulated as a question and the additional content element contains the possible answers in the form of buttons. A further addition is that after selecting an answer, i.e. when the user clicks an answer button, a fitting response is given. The response can be another question (with answer options or open), a remark or be self-disclosing. A special case of this pattern is the yes/no-question. No response is also possible but not preferred.

4.4.3. Open question

The open question pattern is similar to the 'question with answers' pattern. The difference is that instead of presenting possible answers the children can articulate their own answer via a text input field. Again, the answer is usually followed by a reply, in the form of one of the other patterns, but this time no direct link is made with the actual answer. No text processing is applied to the answers at this stage.

4.4.4. Mood matching

Mood matching is a pattern that is applied within all previous patterns when it serves as a response to an added activity. The way the text is formulated and the intensity of the selected gesture is modulated by the mood supplied by the children when they add a new activity. This to facilitate empathic responses.

4.4.5. Ask activity item

Ask activity item is the pattern the avatar uses when it queries the child for activity items. The text component consists of a question asking for one if the activity items. One of the high intensity gestures is always used for this pattern. Finally, the content component fits with the question. For the activity type and the mood items the available options are listed and for the other items the right input fields are displayed (text, text box, date, time, and picture uploading or linking). The activity items are always asked in the same order.

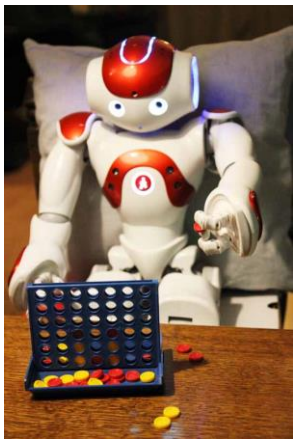


Figure 5: the PAL robot playing connect-4.

4.4.6. Self-disclosure

The self-disclosure pattern is also similar to a remark. The text is usually longer than in a remark or more self-disclosure items are modelled to come after one other. The content of the text is always about the avatar and presented from the perspective of the avatar, e.g. "I like to play football as well". Most of the self-disclosure texts are about what the avatar likes or dislikes, has done that day or still wants to do. The gesture component is fitting with the contents of the text. The additional content component either contains nothing or a picture of the physical robot. The picture always fits with the text. The pictures display the PAL robot undergoing an activity such as playing football, going to the library, playing connect-4 (see), etc.

4.4.7. Encourage

The encourage pattern has the function to remind the children to perform a certain action, such as adding a new goal or completing an existing one, without putting too much pressure on them or coerce them. The content of the text component therefore is non-directive and formulated positively, for example "A goal is a good way to keep track of your progress. Maybe you want to add one?" or "You made some good progress with your daily goal today. Keep it up!". One of the high intensity gestures is used for this pattern. The additional content component is empty.

4.4.8. Praise

The praise pattern has the function to compliment the child with an accomplishment such as completing a goal or adding an activity. The praise pattern is a special case of the remark. It is essentially the same with the only restriction that the content of the text is always formulated as a compliment.

4.5. Effects and measures

The listed functions and interaction design patterns all have goals: supporting the feeling of autonomy, competence and/or relatedness. Supporting those feelings has the goal to stimulate the (intrinsic) motivation of the children to keep using the diary. But how successful are the functions and interaction design patterns in achieving those goals? To establish the success, we need to measure the effects of the functions on motivation and diary use? In this section we present the causal chain of intended effects and introduce measure to capture those effects.

4.5.1. Effects: what do we need to measure?

The intended causal model is displayed in Figure 6. The mechanisms included in myPAL, among which the interactions with the avatar, support the feeling of autonomy, competence and relatedness. The self-determination theory argues that supporting these factors stimulating circumstances are created for the children to be motivated to use the diary. The aim is that the created circumstances will result in a consistent use of the diary with plenty of added content.

MyPAL is built to follow the displayed causal chain. The first step in evaluating myPAL is to measure these factors and see if this causal chain is realized. In Figure 33 all the operationalizations of these factors are displayed.

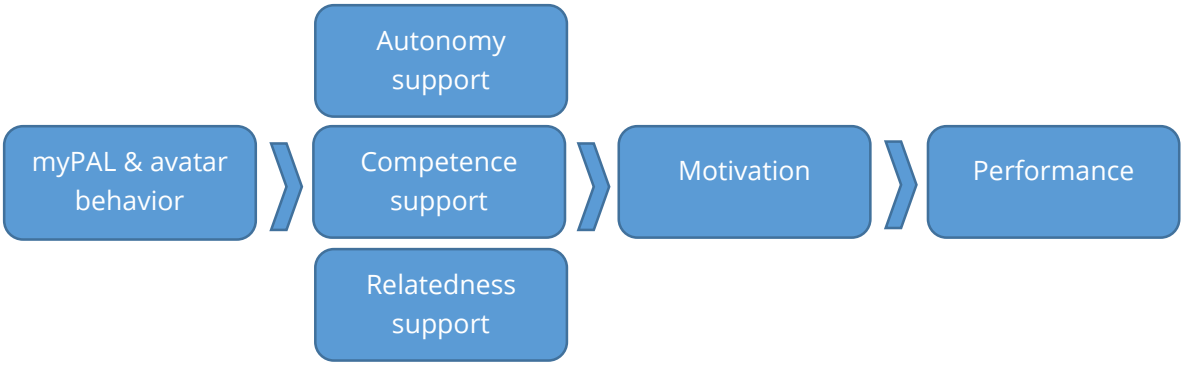


Figure 6: the intended causal model for stimulating a consistent use of myPAL

4.5.2. Measures: how are we going to measure it?

Capturing the effect of myPAL and the avatar is not a straightforward process in the current set-up due to the lack of a control group. The way we still can get some sense of the existence of an effect is by first establishing whether or not the participants observed the behavior of myPAL and the avatar. To reduce the complexity of this step I will primarily focus on the behavior of the avatar in this thesis. Once established how well the participants observed the behavior the effect of the amount of observed behavior on the antecedents of motivation can be tested.

Furthermore, an alternative route is to first determine the attitude of the participants towards the avatar in terms of perceived sociability, appreciation of the avatar behavior and self-assigned importance of the avatar. Research indicates that these factors might modulate the effectiveness of the avatar [16].

Another aspect is measuring motivation. Motivation is a psychological construct and cannot be measured directly. One of the most often used and well validated approaches is to measure motivation via self-reports, e.g. using a questionnaire. However, the validity and value of this technique is questioned, because it is sensitive to a bias from social desirability concerns and it is limited to one's conscious understanding of one's own motivation. It is argued that, implicit cognitive, affective and behavioral measures are preferred [62], [63].

Examples of implicit measures are:

- *Cognitive – Lexical/Object-Decision Task*

  The accessibility or readiness of goal-related concepts is higher when one is more motivated to reach that goal. This can be measured using a lexical-decision task, i.e. determining whether a string of characters is a word or a non-word [62], [64]. Instead of a lexical-decision an object-decision task (using pictures instead of words) is also possible to measure whether concepts are activated in memory [65]. For example, when the goal is to update digital diabetes diary, an image of a paper diary would (when motivated) be faster distinguishable (from a non-image i.e. a 'blob') than of a neutral image e.g. of a ship.

- *Affective - Evaluative Priming Task*

  The degree of which a goal-relevant concept is evaluated positive is an affective approach to measure how motivated a person to complete that goal [62]. One way to measure that is to prime people with goal-related concepts, e.g. "taking a picture of food", and measure

the time it takes to identify positive words, e.g. “great” from non-words versus negative words, e.g. “dirty” [66].

- *Affective - Implicit Association Test (IAT)*

  Another affective approach is the implicit association test (IAT) [62]. Two target concepts, e.g. diary (salient) and ships (non-salient), are linked to evaluation categories, e.g. pleasant/motivating or unpleasant/non-motivating words. These resulting word pairs can either be congruent (diary + motivating vs. ships + non-motivating) or incongruent (diary + non-motivating vs. ships + motivating). The word pairs are displayed left or right on the screen. People have to link a target word to either of the word pairs by pressing the corresponding (left or right) button. When people classify diary or motivating related target words faster in the associated condition compared to the less associated condition they, in this case, associate diary with motivation [67]. This implies a higher level of motivation [62].

- *Affective - subjective experience - Intrinsic Motivation Inventory (IMI)*

  Motivation can also be measured by evaluating the subjective experience [62]. The Intrinsic Motivation Inventory is developed (and validated) to capture the subjective experience by investigating the level of interest, perceived competence, effort, usefulness, felt pressure and tension and perceived choice [68]–[70].

- *Behavioral – choice*

  Presenting people with the choice to do a goal-congruent or incongruent activity is indicative for the motivation to achieve a certain goal [62]. Read, Loewenstein, and Kalyanaraman (1999) use the ratio of goal congruent/incongruence choices as a measure of motivation [71].

Unfortunately, not all the listed methods are available as standardized or validated tests for the purpose of measuring motivation related to keeping a diary. Only the intrinsic motivation inventory had validated items for this purpose. Therefore, in the performed user study the questionnaires that were used to measure motivation and its antecedents were based on validated IMI items. See section 6.2.2 for the exact items and sources.

In the end, motivation shows itself through action. Actions results in behavior that can be detected to subsequently infer motivation. Performance is one of the most often used measures [62]. With myPAL we aim to stimulate children to add plenty of content as consistently as possible. Therefore, to measure the performance of myPAL we use the following measures:

- Amount of added content.
- Consistency of added content.

5. Build prototype: myPAL

In the previous chapter's extensive work has been presented that all leads to the specifications of a cognitive software system – a digital diabetes diary with responsive avatar. A tree-based decision model is used to structure the avatar behavior model. Furthermore, myPAL is developed as a web application based in the Play Framework. Rapid prototyping is used to advance from working prototype to working prototype. In this chapter I will discuss the implementation of the avatar behavior model, the used materials and tools and I will describe the final prototype.

5.1. Avatar behavior model

In Figure 8 the decision model of the avatar behavior is depicted. The avatar is either triggered by a page load or a button press. Currently only the 'add activity' button leads to avatar behavior, namely the avatar will ask the child whether he or she wants to add an activity alone or together. An avatar reasoning engine will traverse the tree after each user action in the web application. The reasoning engine expands decision nodes based on currently available information. A leaf node consists of one or more sets of behaviors together with a probability. Given the probability a behavior set is randomly selected. The behaviors are extracted, loaded and executed by the avatar. In Figure 8 the id's of the actual behaviors are listed. Those id's match to the behaviors listed in Appendix 3.

The 'page' decision node is based on a combination of the child's and avatar's behavior history. In four situations the avatar is triggered. The first situation is when an activity is added. The avatar will formulate a response based on the mood of the child and the activity type (school, sport, meal, other). The second situation is when the child is logged in. The avatar will then greet the child. The third situation is when the child wants to add the activity together. The avatar will ask all information items one by one. The fourth and currently the last situation is when the child accesses the goal page.

The goal page can have one of the following statuses: goal met, goal added, goal active, no goal active. When the child accesses the goal page and he or she just met a goal the avatar will praise the child. When the child accesses the goal page after just adding a goal the avatar will praise the child as well. When no goal is met or added but a goal is active the avatar will cheer the child on to complete the goal. And finally when no goal is active the avatar will recommend the child to add a new goal. The behaviors are differently for daily goals and total goals.

5.2. Materials and tools

MyPAL is fully responsive web application. This means that it is accessible, in a user friendly way, from any device (desktop, laptop, smartphone and tablet) independent of screen size. The web application is developed using the Play Framework[1]. Play is an open-source Java and Scala based framework that includes a lot of libraries which make web development easier. It employs the model-view-controller design pattern separating the concerns for data storage, data representation and user interaction. The core concept works as follows (see Figure 7). When a user tries to access for example 'www.websiteaddress.nl/goals' the route '/goals' will lead to a method call of 'getGoalPage()' in the controller class. The controller class retrieves all the information that

[1] https://www.playframework.com/

needs to be represented on the goal page, e.g. all active goals of that user, by accessing the Goal model class. The controller class adds the retrieved information to a goalPage view Scala template. The Scala view template contains the HTML code of the goal webpage. By rendering the template, the added goal data gets injected into the HTML, e.g. in a table. The controller class returns the rendered HTML goalPage. Play serves the rendered HTML page to the user.

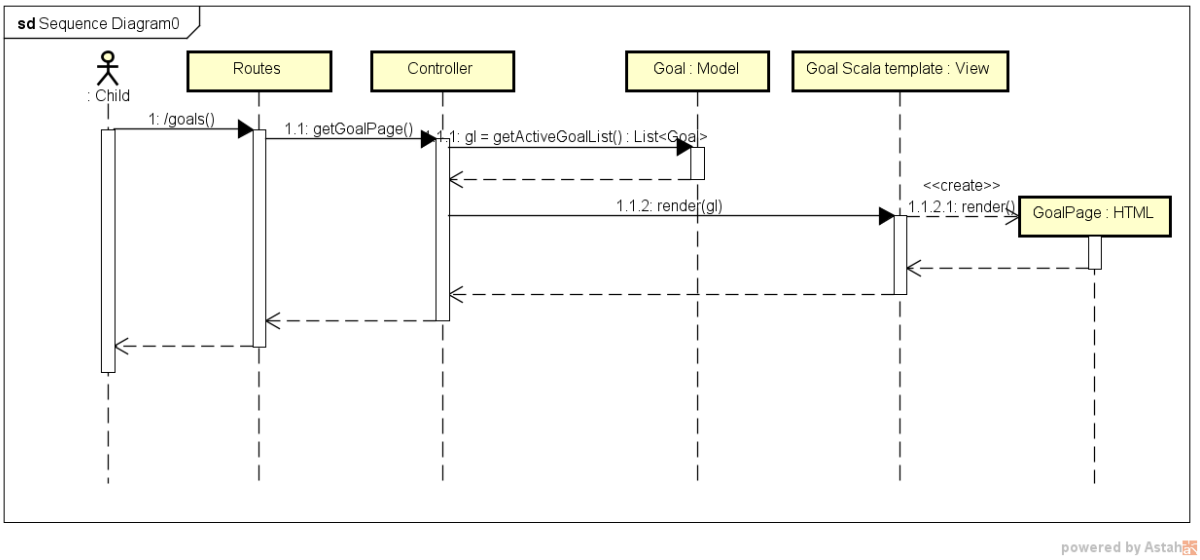


Figure 7: sequence diagram of an example of the core concept of the Play Frameowrk

An H2 Java SQL database is used to store the added content persistently. Ebean, an object-relational mapping (ORM) library, is used to link Java model classes to tables in the database. Creating, reading, updating and deleting model objects in the Java code will automatically lead to the equivalent operations in the database without having to write SQL queries.

The session() method is used to validate the authorization of logged in users. Certain parts of the web applications are only meant for a researcher user. When a user is logged in a session() is created which stores the user name. Every time a webpage is accessed the session is used to retrieve the user information. When the session is empty the user is automatically rerouted to the login page. When the user information is successfully retrieved a check if performed to verify whether the user can have access to the page. The page will only be loaded and displayed when the user has the right access rights.

To not spend too much time on the graphical design of the webpage without giving up on child friendliness and a fully responsive character Bootstrap is used. Bootstrap is an open source graphical framework for rendering webpages using HTML, JavaScript and CSS[2]. By labeling HTML components in the Scala view templates, e.g. a form, Bootstrap knows what to provide with a Bootstrap makeover.

[2] http://www.getbootstrap.com/

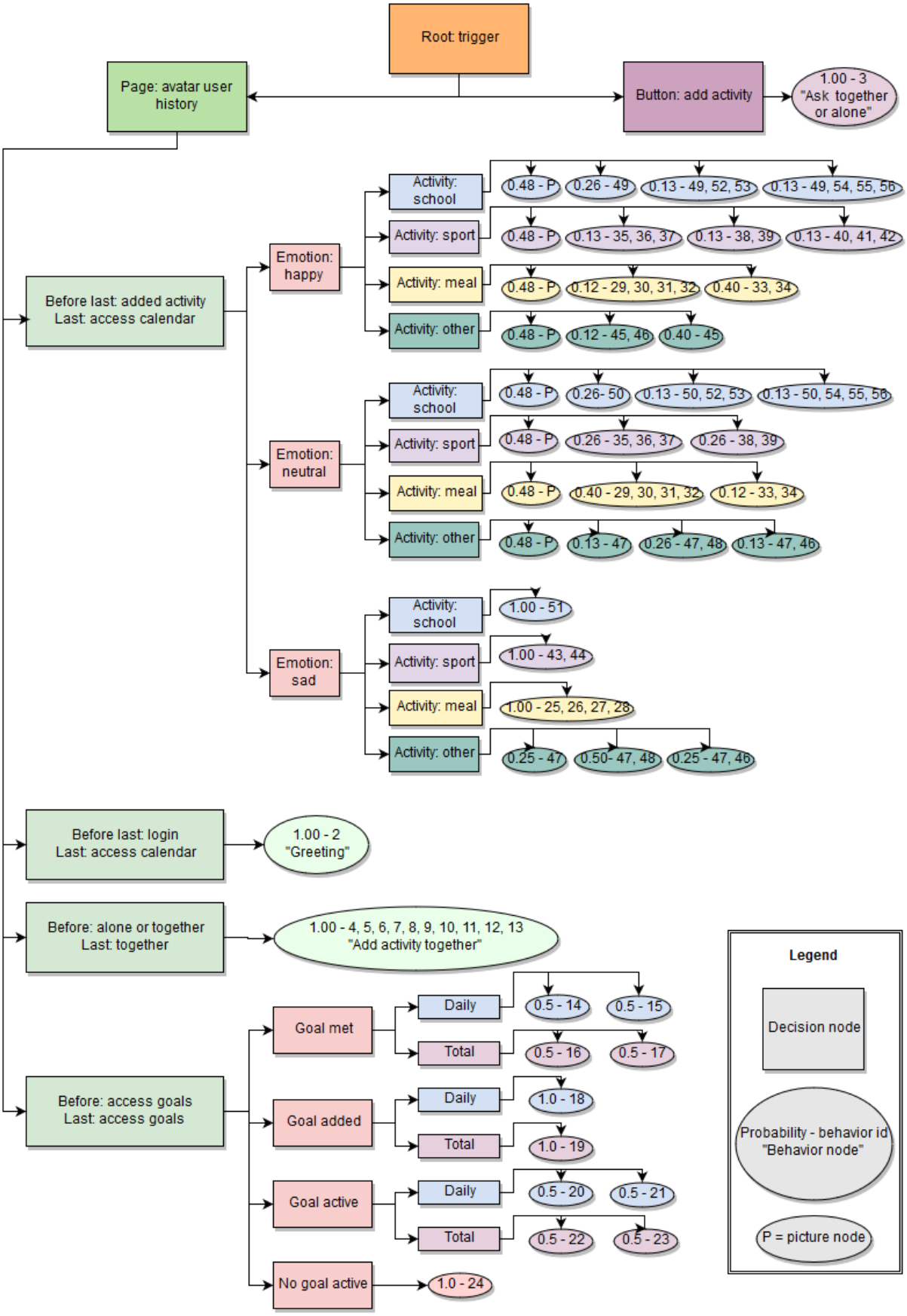

Root: trigger
Page: avatar user history
Button: add activity
1.00 - 3 "Ask together or alone"
Before last: added activity Last: access calendar
Emotion: happy
Activity: school
0.48 - P
0.26 - 49
0.13 - 49, 52, 53
0.13 - 49, 54, 55, 56
Activity: sport
0.48 - P
0.13 - 35, 36, 37
0.13 - 38, 39
0.13 - 40, 41, 42
Activity: meal
0.48 - P
0.12 - 29, 30, 31, 32
0.40 - 33, 34
Activity: other
0.48 - P
0.12 - 45, 46
0.40 - 45
Emotion: neutral
Activity: school
0.48 - P
0.26- 50
0.13 - 50, 52, 53
0.13 - 50, 54, 55, 56
Activity: sport
0.48 - P
0.26 - 35, 36, 37
0.26 - 38, 39
Activity: meal
0.48 - P
0.40 - 29, 30, 31, 32
0.12 - 33, 34
Activity: other
0.48 - P
0.13 - 47
0.26 - 47, 48
0.13 - 47, 46
Emotion: sad
Activity: school
1.00 - 51
Activity: sport
1.00 - 43, 44
Activity: meal
1.00 - 25, 26, 27, 28
Activity: other
0.25 - 47
0.50- 47, 48
0.25 - 47, 46
Before last: login Last: access calendar
1.00 - 2 "Greeting"
Before: alone or together Last: together
1.00 - 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 "Add activity together"
Before: access goals Last: access goals
Goal met
Daily
0.5 - 14
0.5 - 15
Total
0.5 - 16
0.5 - 17
Goal added
Daily
1.0 - 18
Total
1.0 - 19
Goal active
Daily
0.5 - 20
0.5 - 21
Total
0.5 - 22
0.5 - 23
No goal active
1.0 - 24
Legend
Decision node
Probability - behavior id "Behavior node"
P = picture node

Figure 8: avatar behavior model. A reasoning engine traverses the tree by expanding decision nodes (rectangles) given information about the trigger, avatar and user history, user mood, activity type and the status of the goal page. Given a distribution randomly a set of behaviors is selected from the leaf nodes (ellipse).

5.3. Final prototype description

The myPAL system consist of a web application with an embedded responsive avatar. There are two type of users in this version of myPAL: children and researchers (admin). From both user perspectives a description is provided. The description will be in terms of user paths i.e. common interaction the users can have with the system.

5.3.1. MyPAL for children

The diary system contains a browsable calendar where children can create, read, update and delete (CRUD) activities and measurements. Furthermore, it contains a page to set daily goals and reoccurring goals. Daily goals need to be met within 24 hours and reoccurring goals are about doing an action, e.g. adding an activity, for multiple days in a row. The last main functionality is the photo gallery where all the added pictures of a user are shown. Pictures can be added directly to the gallery or be added together with an activity. When pictures are added during the creation of an activity they are linked to that activity. Directly added pictures are unlinked. Unlinked pictures can be linked to an activity and vice versa.

When children open the myPAL website in their browser they will see a login screen (see *Figure 9*). They have to enter their username and password to access their diary.

Figure 9: Login screen

When children are logged in they are rerouted to the calendar page (see *Figure 10*). There they are greeted by Robin the avatar. The 'PAL-blue' bar at the top of the screen can be used to navigate to the other main functionalities such as the calendar, goals, gallery and logging out. The admin button is only present when the user is a researcher and links to the admin pages.

Children can do three things in the calendar. They can browse it by going up or down a day, by clicking the arrows, or select a day manually by clicking the button that shows the current selected day (between the arrows). The second and third thing a child can do is to respectively add an activity or measurement by clicking the corresponding buttons. In the center of the screen the text 'Voeg een activiteit toe.' (add an activity) is only shown when no activities or measurements are added on the selected day. It is a clickable instance as well and redirects to the 'add activity' page.

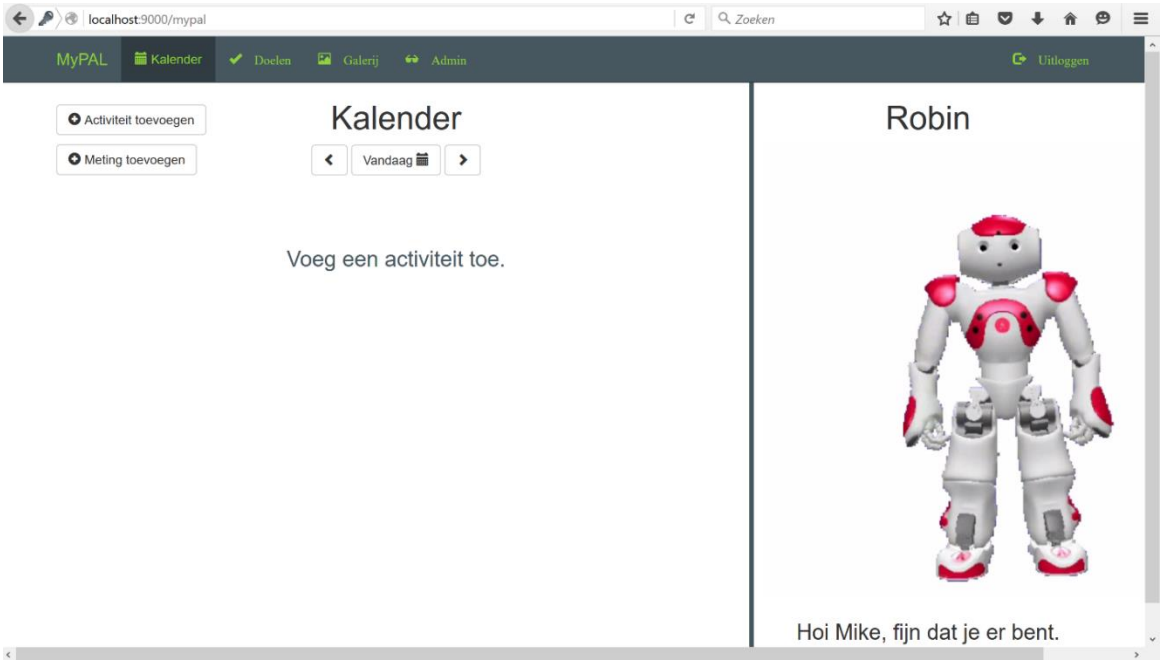


Figure 10: Home screen

When children choose to add an activity the avatar asks them whether they want to add it on their own or together with the avatar (see *Figure 11*).

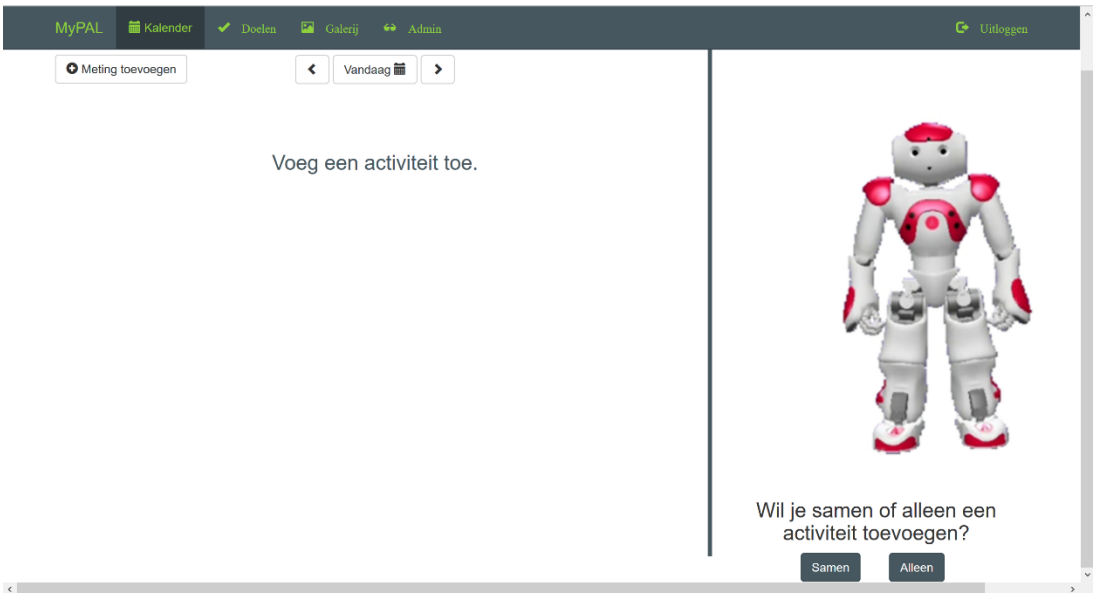


Figure 11: Add activity together or alone?

When the children choose to add it alone they get to fill out a form with all the required information (see *Figure 12*). The only optional step is the adding or linking of a picture. It is allowed that activities have no pictures. When children want to add an activity together the avatar will ask them the same questions step by step (see *Figure 13*).

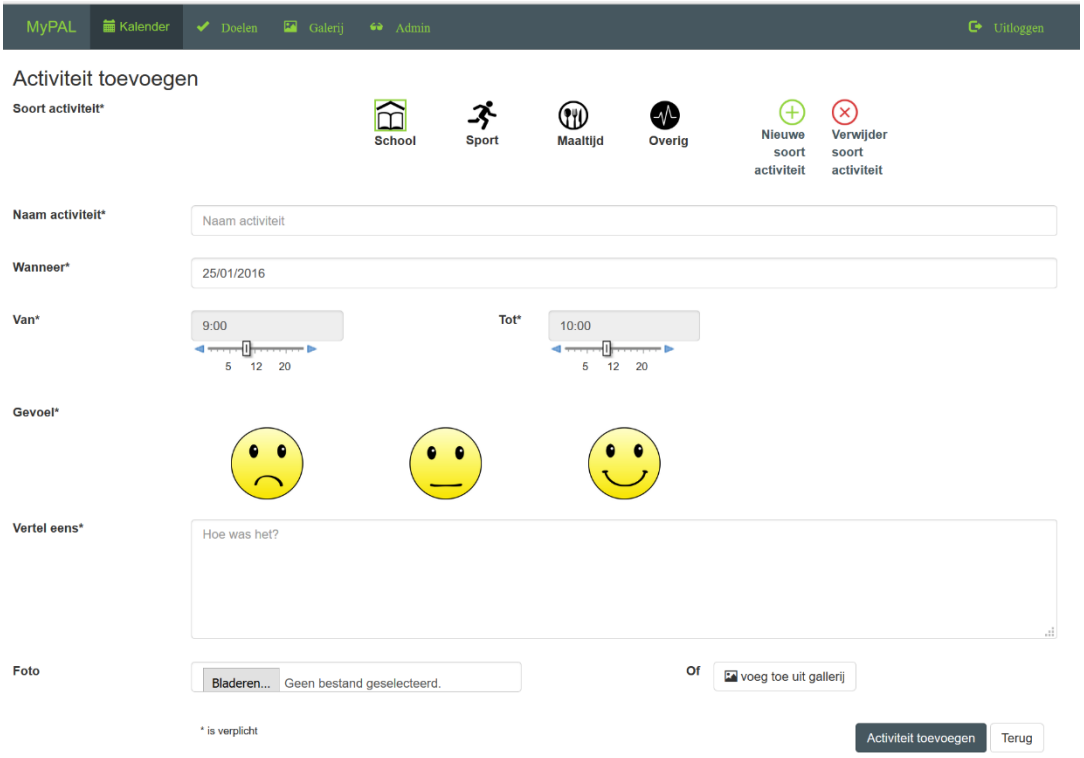


Figure 12: add activity form

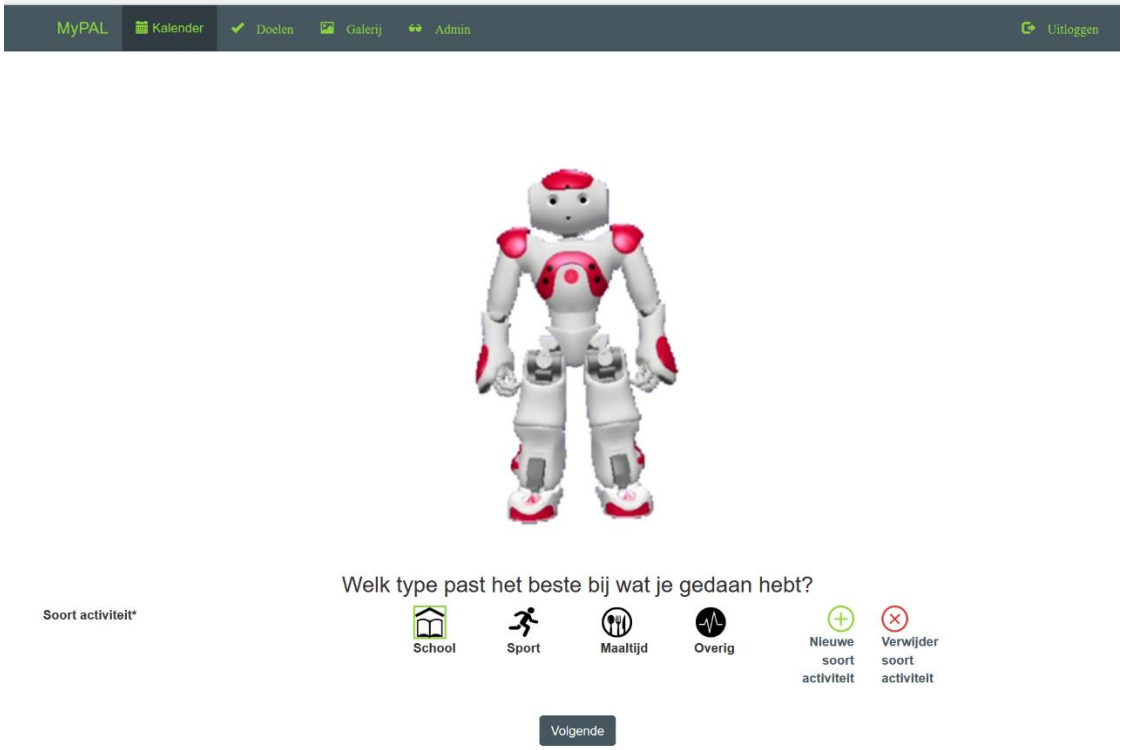


Figure 13: add activity together

Children are required to add the type of the activity, e.g. school, meal, sport, a title, date and time, the emotion they associate with that particular activity and a description. As said adding a picture

is optional. Based on the type of the activity and the reported emotion the avatar will respond to the added activity (see *Figure 14*). The avatar can also share a picture of one of his adventures (see *Figure 15*).

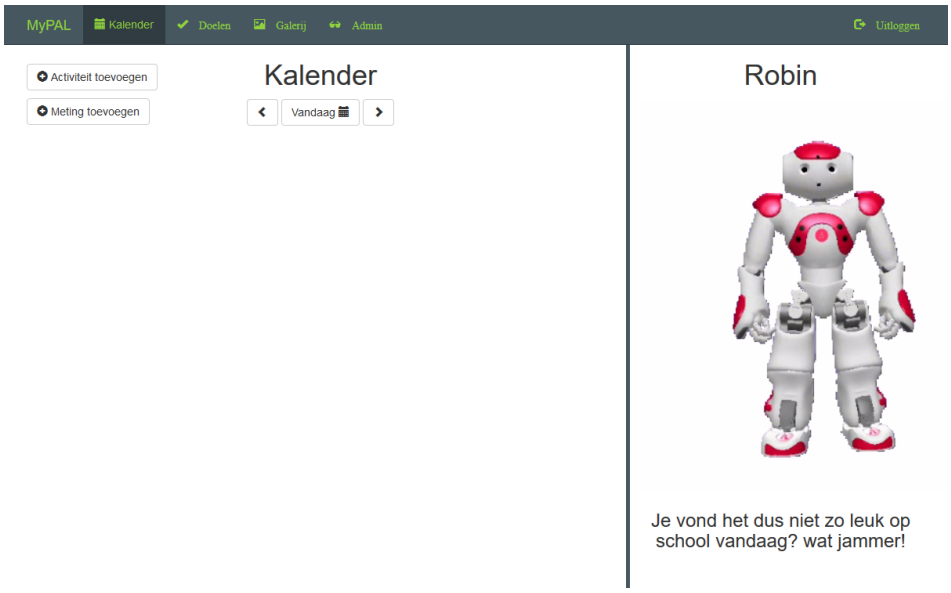


Figure 14: avatar responds to unhappy school activity

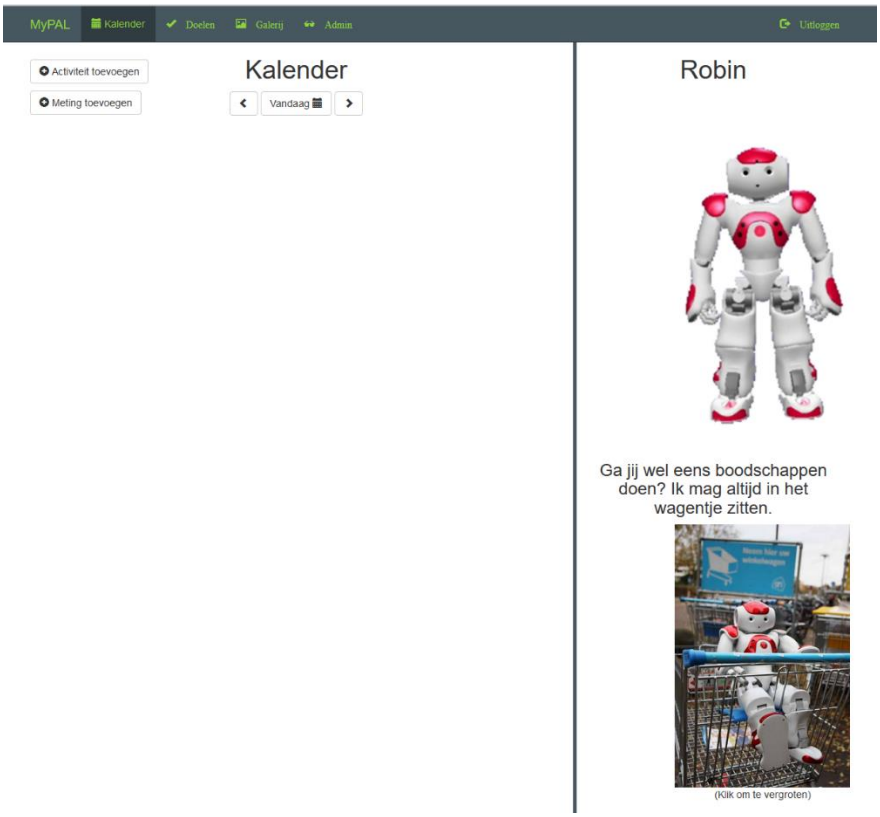


Figure 15: avatar sharing a picture of his own

After activities are added they appear in the calendar page. Activities can be selected by clicking the associated bar in the display. This will bring up more information about that activity as shown

in *Figure 16*. By clicking on the cross and pencil button the activity can respectably be deleted or updated.

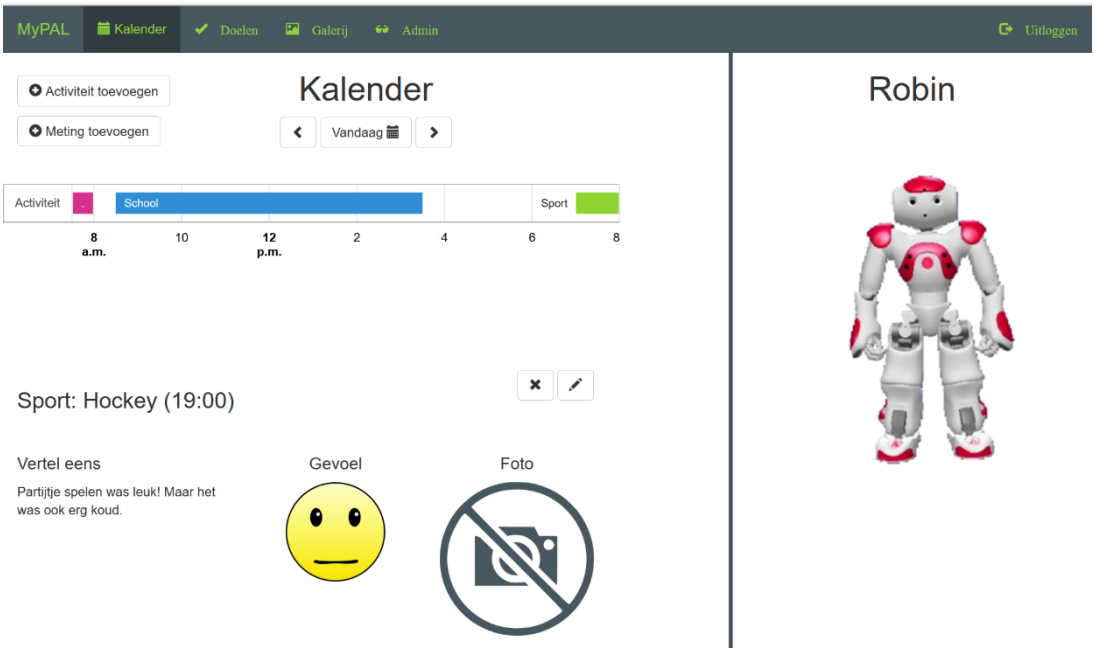


Figure 16: viewing an activity

Children can add two types of measurements: blood glucose values or (additional) insulin intake. Measurements can only be added through a form (see *Figure 17*).

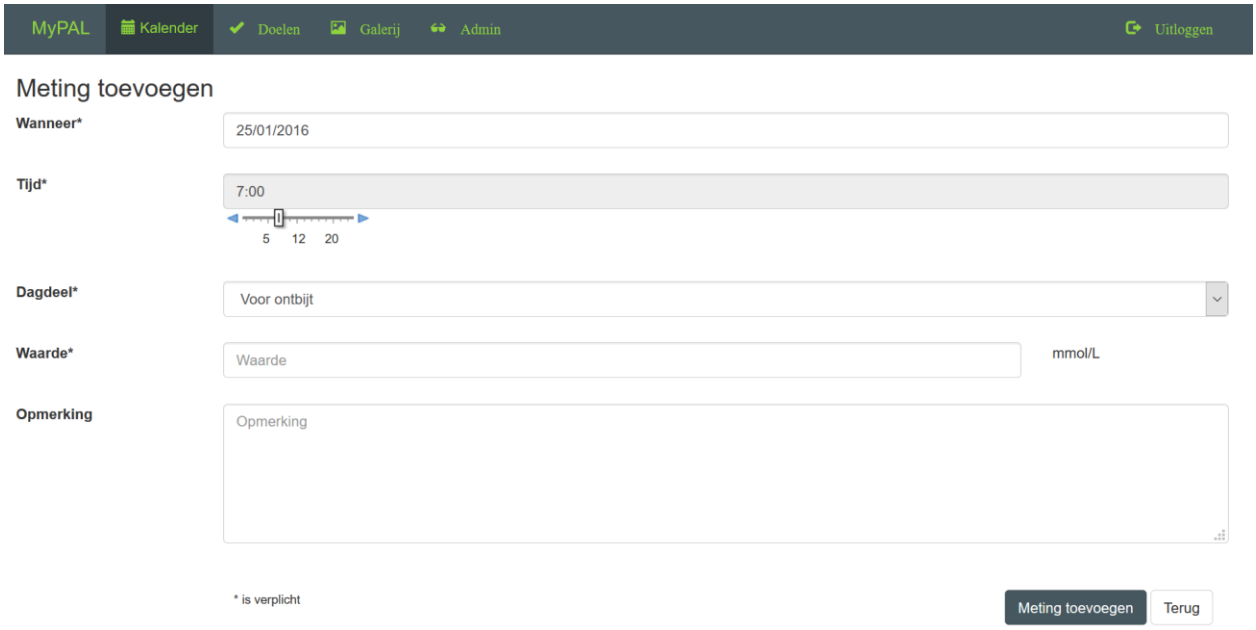


Figure 17: adding measurement

From the gallery the children can see their linked and unlinked pictures (see *Figure 18*).

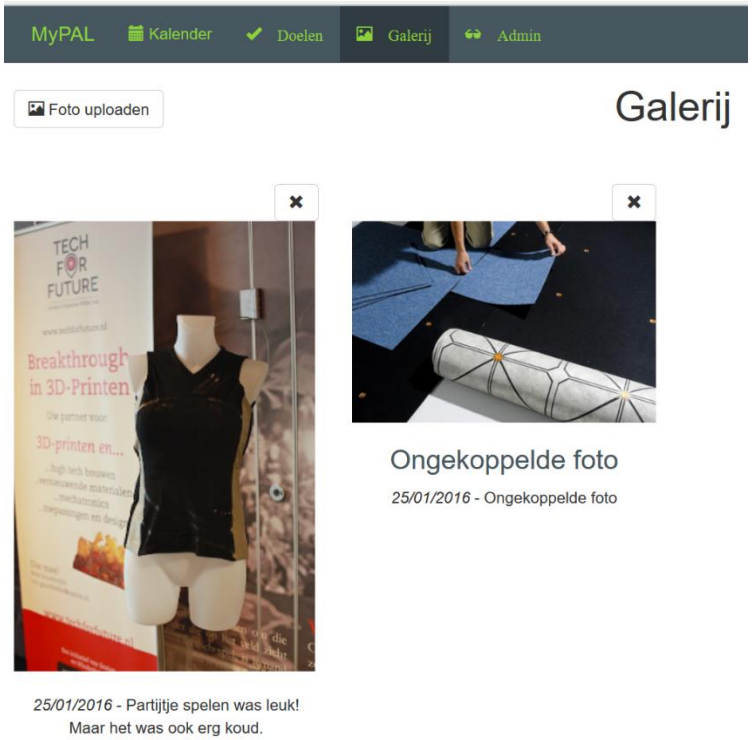


Figure 18: photo gallery

When a child navigates to the goal page and there are no goals active the child is encouraged to add a new goal (see *Figure 19*). When a goal is added but still in progress the children are encouraged or given hints on how to complete the goal.

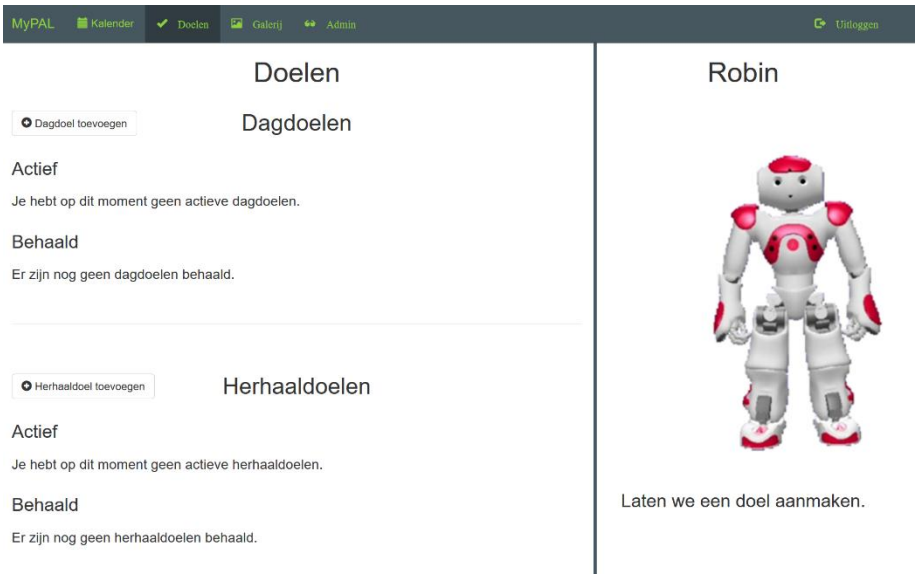


Figure 19: goal page

Children can add daily goals or reoccurring goals. Daily goals need to be met within 24 hours. Children make an agreement with themselves to either add *N* activities, measurements, pictures, or activities from the day before (see *Figure 20*). The value of *N* can also be selected by the children. *N* lies between 1 and 10.

Figure 20: Add daily goal

Reoccurring goals are again agreements that children make with themselves (see *Figure 22*). But now they agree to log in once or add at least one activity, measurement, or picture for *M* days in a row. Children can give a value to *M* between 2 and 7 days.

Figure 21: add reoccurring goal

Children can view the progress of their goals on the central goal page (see *Figure 22*). The avatar will give encouragements or tips to meet the goals in time. The avatar will also praise the child when a goal is met. Furthermore, all the met goals are collected in this view (the goals with a check mark) as well as additional information about a goal (when selected).

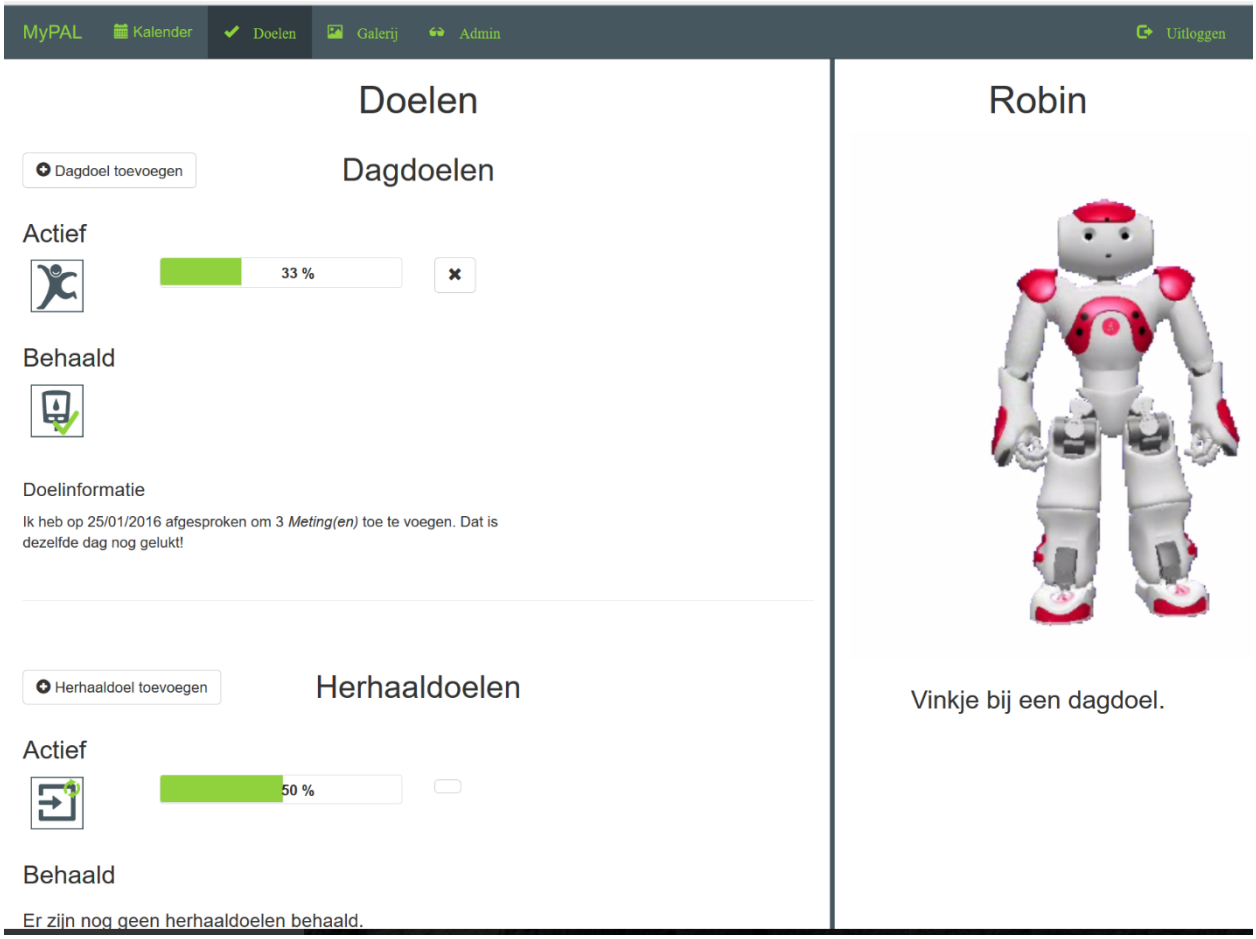


Figure 22: The progress of the active goals can be viewed

5.3.2. MyPAL for researchers

A researcher is able to create, read, update and delete users. A researcher can see the user content, behavior logs of users and the avatar, and the demographical information per user. Furthermore, the researcher is able to create, read and delete gestures and behaviors. An avatar gesture is an animation (video clip) or image of a red Nao (PAL robot). An avatar behavior is a dialogue element (audio + text) linked with a gesture.

When a researcher logs in he or she is redirected to the researcher dashboard page (see Figure 23). The researcher can see how many users are registered and how many of them are currently online. Furthermore, counts are displayed for the added activities, measurements, available avatar behaviors and gestures, and user behavior log items. Also the most recent added activities are displayed and the option is given to download the log files and added content.

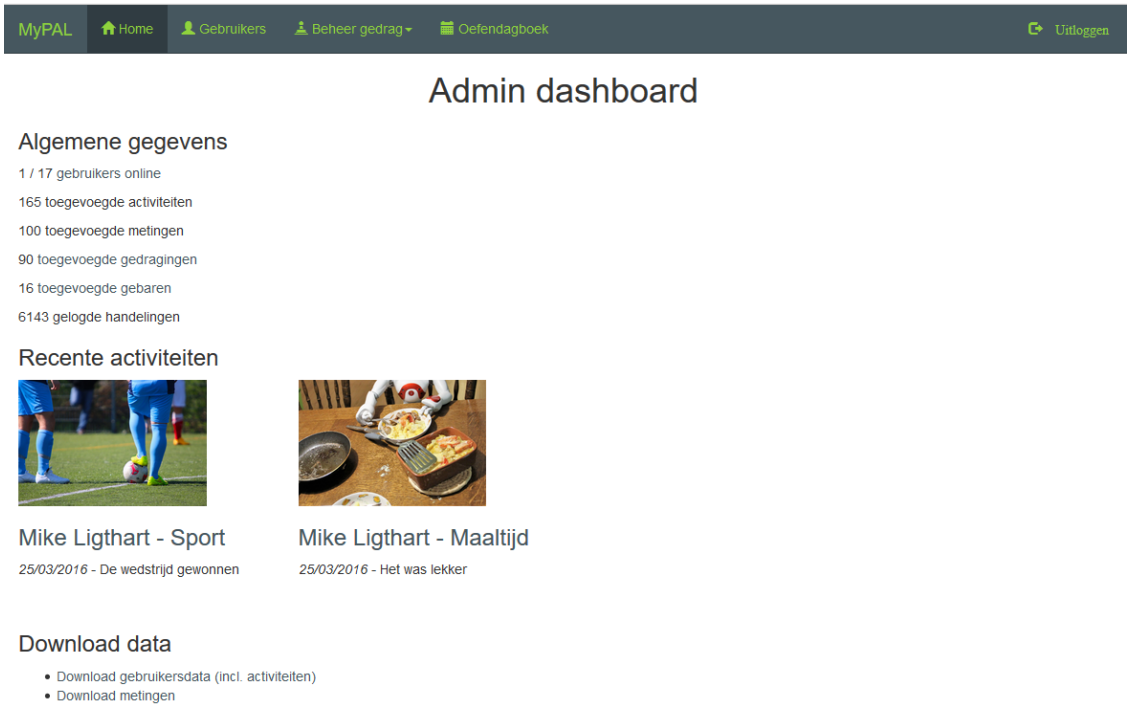


Figure 23: Admin dashboard containing counts of all the added activities, measurements, etc., an overview of the recent added activities, and the option to download the all the added user data.

When the researcher selects the users ('Gebruikers') tab from the main menu the user overview page is loaded displaying all the registered users (see Figure 24). The page also contains a form where a new user can be added. To create a new user their username, first and last name, birthdate, user type, (optional) gluconlineID and password need to be filled in. The same account information is displayed below the registration form of all the entered users. Instead of the birthdate the age of the user is displayed together with the time and date when they were last active. Furthermore, for each user an option is shown to edit, delete or to 'view' them. That last option loads a new page where all the detailed user specific information is displayed (see Figure 25, Figure 26 & Figure 27). The user information page shows again the account information together with an overview of all the added activities and measurements and the logged user and avatar behaviors.

MyPAL | Home | Gebruikers | Beheer gedrag | Oefendagboek | Uitloggen

Gebruikers

Voeg gebruiker toe

Gebruikersnaam*	Gebruikersnaam
Voornaam*	Voornaam
Achternaam*	Achternaam
Geboortedatum*	01/01/2005
Gebruikerstype*	Kind
Wachtwoord*	Wachtwoord
GluconlineID	GluconlineID

Voeg gebruiker toe

Overzicht

10 resultaten weergeven — Zoeken:

Gebruikersnaam	Voornaam	Achternaam	Leeftijd	Gebruikerstype	GluconlineID	Laatst actief	Opties
[redacted]	[redacted]	[redacted]	10	Kind		01/06/2016 13:29:51	
[redacted]	[redacted]	[redacted]	11	Admin		02/18/2016 10:25:06	
[redacted]	[redacted]	[redacted]	11	Kind		01/12/2016 19:16:02	
[redacted]	[redacted]	[redacted]	9	Kind		12/14/2015 17:19:48	
[redacted]	[redacted]	[redacted]	10	Kind		12/13/2015 11:09:01	
[redacted]	[redacted]	[redacted]	10	Kind		02/26/2016 15:35:39	

Figure 24: Top: form to create new users. Bottom: overview of all registered users with the option to view, delete or edit them.

Mike Ligthart

Ga terug naar het overzicht

mike.ligthart

Geboortedatum: 26/11/2015 (0)
GluconlineID: 067835880
Laatst actief: 03/25/2016 12:24:22

Overzicht activiteiten

10 resultaten weergeven — Zoeken:

id	Wanneer	Starttijd	Eindtijd	Soort activiteit	Naam activiteit	Gevoel	Bevat foto	Vertel eens
265	25/03/2016	17:00	17:30	Maaltijd	Avondeten	HAPPY	Ja	Het was lekker
264	25/03/2016	18:45	20:15	Sport	Voetbal	HAPPY	Ja	De wedstrijd gewonnen
65	27/11/2015	09:00	10:00	School	asd	NEUTRAL	Nee	asd

1 tot 3 van 3 resultaten — Vorige | 1 | Volgende

Overzicht metingen

10 resultaten weergeven — Zoeken:

id	Wanneer	Tijd	Soort	Waarde	Eenheid	Opmerking
Geen resultaten aanwezig in de tabel						

Geen resultaten om weer te geven — Vorige | Volgende

Figure 25: Information on a specific user part 1/3. Is shows general information and the added activities and measurements.

Logboek

10 resultaten weergeven Zoeken:

Tijd	Log
30-11-2015 16:34:59	LOGIN
29-11-2015 20:48:10	LOGOFF
29-11-2015 20:45:02	LOGIN
29-11-2015 20:44:50	LOGOFF
29-11-2015 20:44:45	ACCESSCALENDAR
29-11-2015 20:44:44	ACCESSGOALS
29-11-2015 20:44:44	ACCESSGOALS
29-11-2015 20:44:42	ACCESSCALENDAR
29-11-2015 20:44:41	ACCESSCALENDAR
29-11-2015 20:44:39	DELETEACTIVITY

1 tot 10 van 974 resultaten Vorige 1 2 3 4 5 ... 98 Volgende

Figure 26: Information on a specific user part 2/3 shows an overview of all the logged user behavior.

Avatargedrag

10 resultaten weergeven Zoeken:

Tijd	Gedrag
29-11-2015 20:44:44	REACT_GOAL_NOT_ACTIVE
29-11-2015 20:44:30	REACT_NEUTRAL_OTHER
29-11-2015 20:44:19	ASK_TOGETHER_OR_SELF
29-11-2015 20:44:15	GREETING
28-11-2015 22:12:49	REACT_GOAL_ACTIVE
28-11-2015 22:12:15	ASK_TOGETHER_OR_SELF
28-11-2015 22:11:58	ASK_TOGETHER_OR_SELF
28-11-2015 22:11:51	GREETING
28-11-2015 13:12:33	GREETING
28-11-2015 13:11:48	REACT_GOAL_ADDED

1 tot 10 van 222 resultaten Vorige 1 2 3 4 5 ... 23 Volgende

Figure 27: Information on a specific user part 3/3 shows an overview of all the logged avatar behavior

When the researcher selects the behavior management ('Beheer gedrag') tab in the main menu he or she can select to manage the avatar behaviors or the avatar gestures. When 'manage gestures' is selected the researcher can remove or add new gestures. A new gesture can be created by screen capturing the virtual Nao emulator present in Aldebaran's Choregraph (a graphical user interface to program the Nao). A screen capture can be either a picture (.png) or movie (.mp4) for respectively a still or moving gesture. The screen captured gesture can be uploaded accompanied by a name on the gesture management page.

When the 'manage avatar behaviors' option is selected an overview of all available behaviors is displayed containing a gesture, the text that will be uttered by the avatar and an HTML element (see Figure 28). An HTML element may contain several buttons, a text field or a picture of Robin (the robot version). The text can be marked black, red, green, and yellow. When the text is marked black a speech audio file containing the text is available. The audio file needs to be generated and

manually added to the audio source folder in the current myPAL version. The audio is generated by saving the text-to-speech result of the speech synthesizer present in the physical Nao. When the text is marked red the audio file is not available. It is possible to add certain variable elements to the text, e.g. the name of the activity added by the user. A variable is only presented in the text not in the speech. In the speech more general references are made such as 'it', 'the activity', et cetera. Variables are added to the text by surrounding them with the '#' symbol. All items between that symbol are validated. When a variable is accepted it is marked green. When no such variable exists it is marked yellow. This non-existing variable will be left out of both the text and the audio. Furthermore, each text element can be made up out of several versions. Each version has the same semantic value but differs in word choice or syntax.

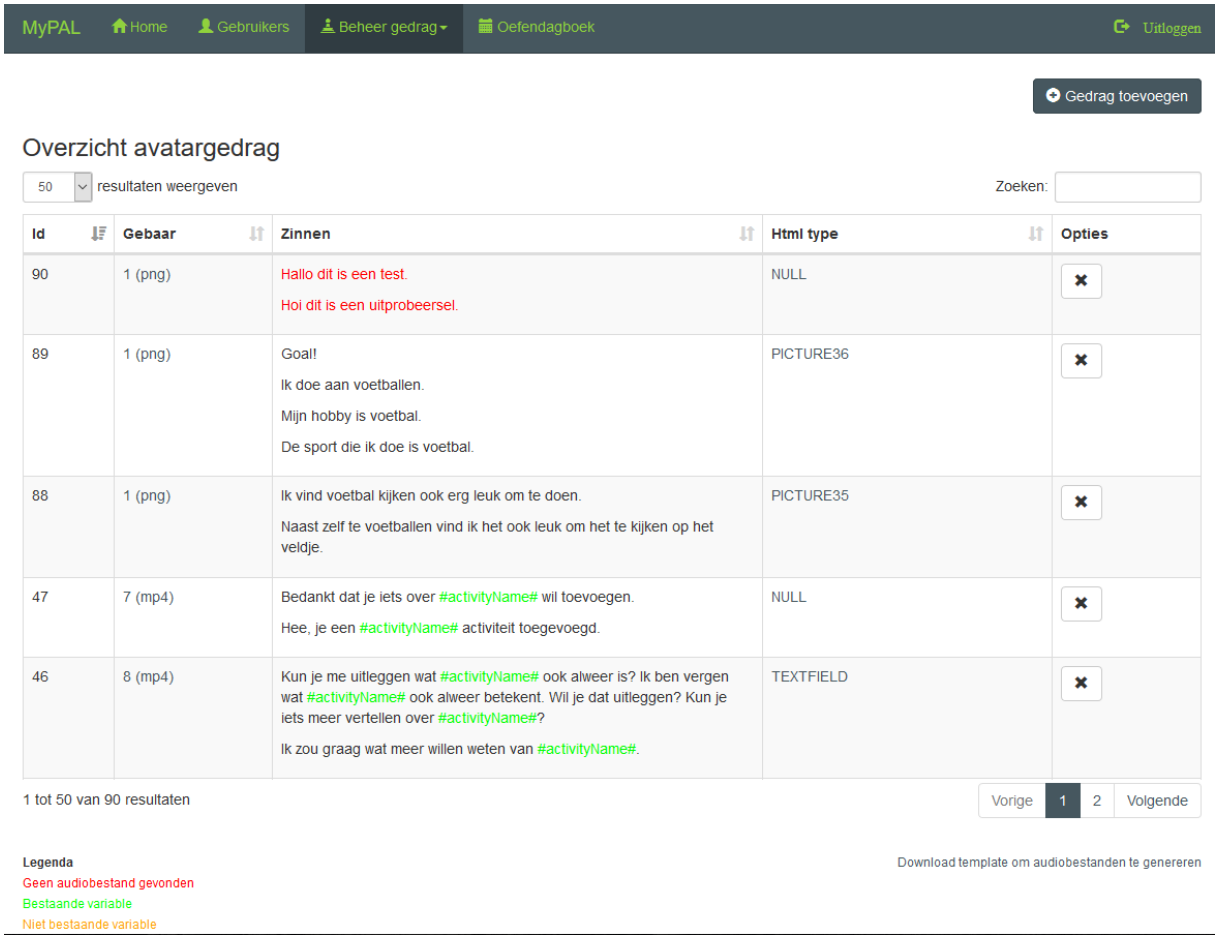


Figure 28 overview page of all avaiable avatar behaviors. A behavior contains an ID, a reference to a gesture, the text and all its variants, an HTML element. A behavior cannot be editted in the current version of myPAL, only added or removed.

From the avatar behavior overview page researchers can navigate to behavior creation page (Figure 29). To create a new behavior, the researcher needs to select a gesture, add the text the avatar should say and select an HTML-element.

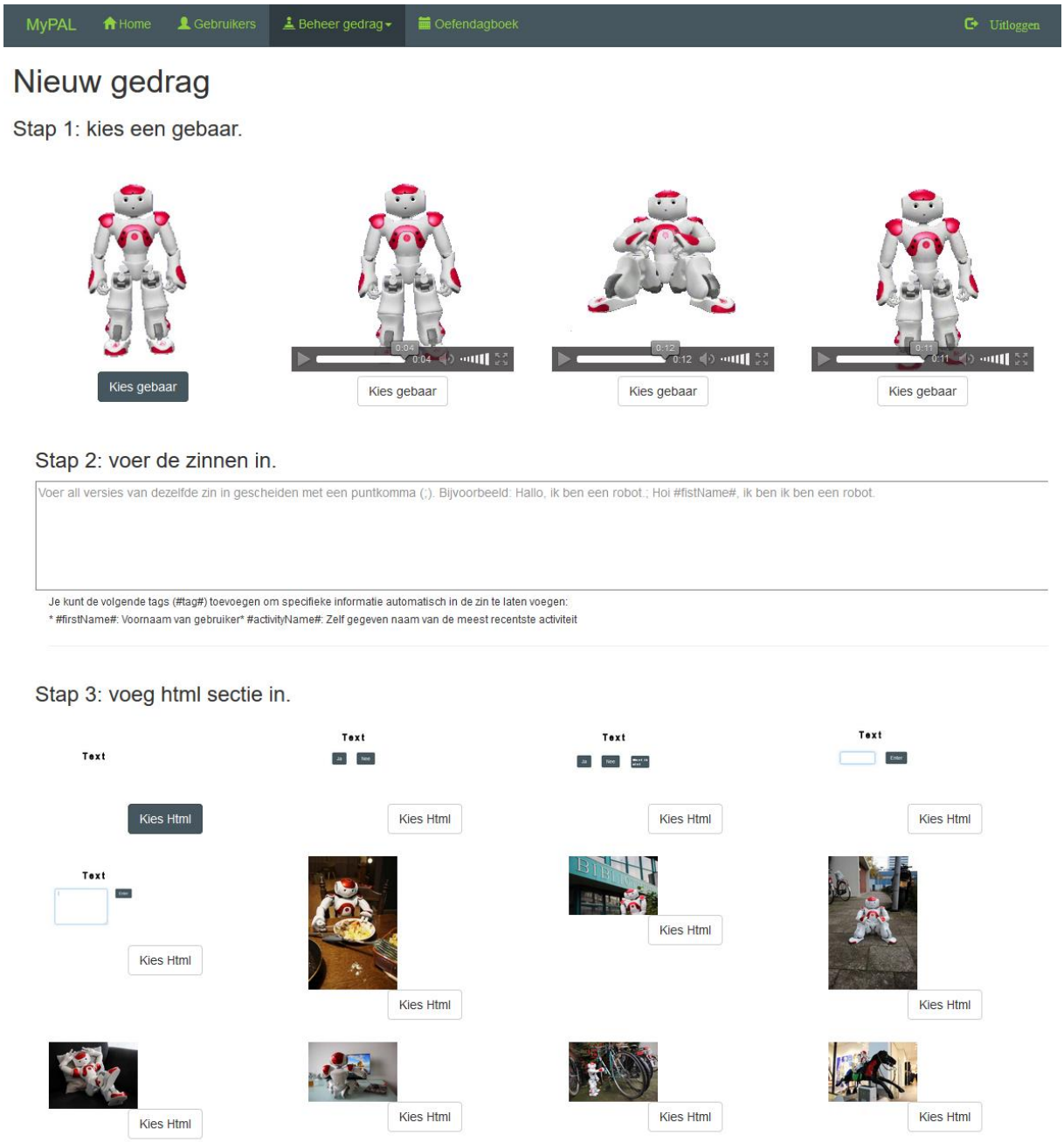


Figure 29: creation page for new behavior. The researcher needs to pick a gesture, enter the text and all its variants and select an HTML element.

6. Evaluation

In the previous sections the process of designing and building myPAL, a digital diabetes diary where the avatar autonomously employs multiple factors to motivate the children for a longer period of time, is discussed. In this section process of evaluating myPAL is described. The evaluation process consists of two steps. The first step was a pilot study at a holiday camp for children with diabetes. The aim was to evaluate the interface and the core functionalities of myPAL and question the children about their use of pictures in their daily life and their preferences for interacting with an avatar. The aim of the pilot study was to identify points for improvements. The second step was a three-week user study with a fully functional myPAL system operating in a natural setting. The goal of the user study was to answer the research questions.

6.1. Pilot study at holiday camp

From 20 till 23 October 2015 the Robots and Heroes holiday camp took place in the rural area of Ommen in the Netherlands. The camp was organized by the children's department of the Dutch Diabetes Association. 21 children between 8 and 11 with diabetes type I and almost an equal amount of staff members and researchers where housed in a cabin in the middle of the woods. The activities ranged from playing games outside to doing chores around the cabin.

A number of activities were organized by researchers from TNO to identify requirements for the PAL system and to test the first prototypes. The activities related to identifying requirements were in the form of creative sessions such as taking pictures, drawing, writing or theater play. The tested PAL prototypes were a sorting game and a quiz that children could play together with the physical robot and an early version of the myPAL application. Only the calendar and gallery sections, without an avatar, were available. A separate interaction with the avatar was used for this evaluation. In this section the evaluation of the myPAL prototype is discussed.

6.1.1. Design and measures

The goals for this evaluation session are to identify points of improvement for the myPAL system, to investigate the attitudes of the children towards the avatar and to collect information about the previous experiences with a diabetes diary and how the children deal with taking photographs. All the children interacted with the avatar and with myPAL in groups and under the same conditions. To achieve the goals, the following elements are collected:

- Answers to avatar questions. The avatar asks the following during his introduction:
    - name of the children;
    - age;
    - whether they have ever used a diabetes diary;
    - If so, how they liked it and if not, why not and
    - whether they would like to keep a diary together with the avatar.
- Behavioral observations during the assignment.
- Questions of the semi-structured interview:
    - Usability-1: what did you thought of the diary. Was it difficult to use it?
    - Usability-2: if you could name one improvement point (each) what would it be and why?

- Avatar-1: what would you think of the avatar when he would reply to what you enter in the diary?
- Avatar-2: what would you want him to say/discuss with him?
- Pictures-1: do you take pictures of the things you do in a day?
- Pictures-2: how do you take these pictures?
- Pictures-3: what do you do with does pictures?
- Pictures-4: would you share the pictures with the avatar?
- Pictures-5: would you like to see pictures of what the avatar does in a day as well?

6.1.2. Participants

21 participants (age range: 8 – 11, median: 9; girls: 8, boys: 13), all diagnosed with diabetes type I, evaluated the myPAL system in pairs or trio's. None of the participants ever used a diabetes diary before nor had they any prior experience with an avatar.

6.1.3. Materials and experimental set-up

The myPAL web application was hosted locally on a standard Dell laptop with internal speakers. The interaction with the myPAL system consisted out of three parts: an introduction with the avatar, trying out the calendar and an interview. The avatar was a simulation of the Nao robot. The introduction consisted of a scripted sequence of explanation or question elements. These elements consisted of an image of the Nao and a pre-recorded speech audio file. The image of the avatar was displayed on the screen together with subtitles of the speech and if needed input elements for the children to answer the questions (see Figure 30). The speech files were generated by the speech synthesizer available in the Nao robot.

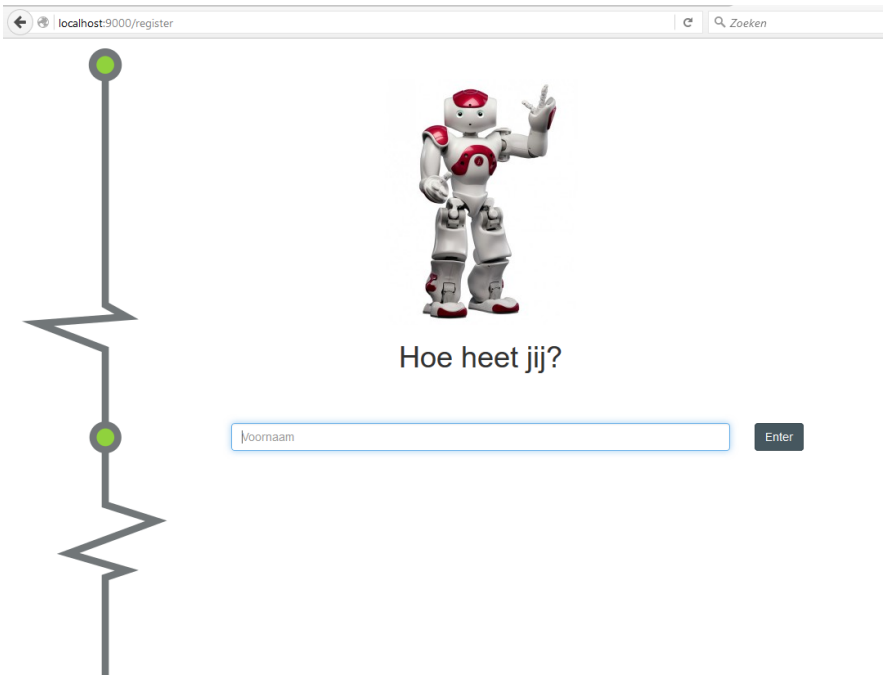


Figure 30: The avatar asking the name of participant. This was part of the introduction of the avatar during the pilot at the holiday camp.

The participants were seated right in front of the laptop so that they could control the mouse and the keyboard while exploring myPAL. The experiment leader was sitting next to the group so that

he could see what was going on and could take over the control of the mouse if he needed to. The interview was recorded using the Smart Voice Recorder app (version 1.7.1), created by Andrew Kovalev, on a Samsung Galaxy S5 smartphone running Android 5.0.

6.1.4. Procedure

MyPAL was part of several prototypes that were evaluated at the camp. The participants were divided in groups of two or three. The groups cycled through a number of stations. At each station a different prototype, e.g. quiz with the robot, was being tested by the children. Per station the children had 15 minutes after which they continued to the next station. myPAL was one of the stations.

Once arrived at the myPAL station the children were instructed to sit directly in front of the laptop that was running the myPAL web application. The experiment leader explained that myPAL was an online diabetes diary that included an avatar. After the explanation the avatar introduced itself and asked the participants for their names, ages, prior experiences with diabetes diaries and their thought about using a diary with an avatar in it. After the questions the avatar explained to the participants that the diary was not finished and that he wanted their help to improve it. From there the experiment leader took over again.

The experiment leader gave a short introduction of the functions available in myPAL at that time, namely the photo gallery and the calendar in which they could add activities and measurements. After the introduction the participants were given the assignment to add a sport they played recently to the calendar without further instructions on how to do it. All the group members took turns to do this once and they were allowed to help each other. While the participants were trying to add the activity the experiment leader observed and wrote down interesting behavior. Furthermore, the discussion within the groups proved to be insightful with regard to the assumptions and experiences the participants had. The experiment leader was allowed to give hints when the participants were unable to complete certain steps in a reasonable time.

The final part of the pilot experiment was a semi-structured interview where the participants were asked to evaluate myPAL and provide insight in whether, what for and how often they take pictures. The list of questions that was used is discussed in section 6.1.1. The experiment leader could ask an unlisted follow-up question to get the participants to clarify, exemplify or deepen their original answer. The questions were directed at individual participants while they were in the group. Sometimes each group member answered the question one by one and sometimes a questions was answered by one and confirmed by the rest. The interviews were recorded on audio.

6.1.5. Results

Two types of results are discussed in this section. A summary of the behavioral observations of the participants are presented together with the answers given during the semi-structured interviews.

Behavioral observations

While interacting with the myPAL prototype without an explanation of the functions all the participants proved to be well able to complete the given assignment of adding an activity. They could easily locate all the necessary functionalities, such as opening the calendar and accessing the right form in reasonable time. There was one element that, in half of the cases, caused confusion. When no activities were added to the calendar a text would appear in the middle of the

page inviting the user to add an activity. This caused some of the participants to click the text thinking it would allow them to add an activity.

The form for adding activities appeared to overwhelm the participants at their first glance. However, most participants started to fill in the input fields discovering that the asked information was in fact basic and familiar. A field that was structurally overlooked was the 'activity type selection' step at the top of the form. As a result, the activity got assigned a wrong type. The input field that was considered the hardest to complete was selecting the time. Not all participants were well practiced with the digital time notation. Furthermore, selecting the time from a drop down menu took the participants quite some effort and time.

The functions of the edit (pencil icon) and remove (cross icon) buttons were not as clear to all participants. Only the icons were displayed on the buttons. When participants were asked to edit or remove an activity they were trying different strategies e.g. adding a complete new activity. After a while most figured out which button to push. Some needed a hint. Once they understood the link between the button icon and the function they considered the relationship as evident.

Interview

The result of the audio recordings of the interview sessions are summarized in this section. One group arrived too late and ran out of time before the interview could start. Per (group of) question(s) a summary is given of the answers.

Usability-1: all the participants agreed that myPAL was easy to use and easy to learn. When asked to think out loud about the consequences of using myPAL for everyday most participants anticipated that using it would become boring after a while. In half of the groups a participant made a remark about how cumbersome filling out the various forms were. These remarks were not always directly directed at the forms themselves. For example, a participant informed about the possibility to automatically load values or other information into the forms. One group noted that they needed to click around once first before being confident enough to complete the assignment. Another group reflected that although myPAL was easy, a paper diary would be even easier.

Usability-2: the participants needed to be stimulated to come up with improvements. The following points were mentioned by at least three groups:

- Improve the adding and updating functionalities of the time.
- Add a tutorial where all elements are introduced properly.

The following points were mentioned by two groups:

- The diary must be able to automatically add parts of the information e.g. blood glucose values.
- Add more default activity types and add default pictures that can be added to an activity.
- Add all the required input stepwise instead of at once to prevent the form to be perceived as overwhelming at the first glance.

The rest of the points were mentioned by only one group:

- Add more color.

- Add the functionality to add future activities and receive reminders for those activities.
- Add funny animations of the avatar.
- Add the functionality to chat with the avatar.
- Add the functionality to give the avatar assignments such as “go stand on your head”.

Avatar-1: all the participants reacted positively to an avatar that would respond to their diary entries. The participants agreed that an avatar would increase their interest of using the diary longer.

Avatar-2: The three most mentioned topics to discuss with the avatar were diabetes care, going over the day and unrelated fun. The participants were especially interested in advise about their diabetes care. For example, help with calculating the bolus. Furthermore, the participants would appreciate it when the avatar asks them about things they have done that day. For example, “how was school?”. Finally, the participants mentioned that they would like the avatar to say and do funny and clumsy stuff.

Pictures-1 to pictures-3: the participants varied on how often they took pictures. This ranged between ‘not really’ to ‘very often’. All the participants who took pictures regularly did this with a smartphone. Furthermore, the pictures were shared in all cased and mostly with parents and friends. WhatsApp was mentioned by most of the participants as the go to sharing medium. One participant stated that she would not share it via social media but just showed them on her tablet.

Picture-4 and picture-5: none of the participants would mind to share pictures with the avatar. Most indicated they would like to share. When asked if the participants would like the avatar to share pictures with them all participants reacted enthusiastically. Some groups mentioned that they hoped that the avatar would share funny pictures.

6.1.6. Discussion

The general tendency from both the behavioral observations as the interviews is that participants find myPAL easy to use and easy to learn. This means that most of myPAL can be used *as is*. Furthermore, a number of leads for improvements have been identified. In this section these proposed improvements are discussed. For each pointer a motivation is given why or why not it was implemented. From the behavioral observations the following improvements are proposed:

1. *Transform the “Add an activity” message, shown on an empty calendar page, into a link to the ‘add activity’ form.*
When no activities are added to the calendar a message is shown instead of the calendar. It encourages the participant to add an activity. Participants tried to click on this message in order to add an activity. This was not possible. Because the participant intuitively thought the “add an activity” message would be a link and changing the text into a link was a minor change this improvement was implemented.

2. *Make the ‘add activity’ form less overwhelming at first glance.*
Although participants thought the form was easy to understand after the first use it looked intimidating at first glance to some. Two improvements have been implemented to prevent this. First of all, all the labels for the input fields are reformulated to a question-like format. For example, the label ‘date’ is changed to ‘when?’ and ‘description’ is changed to ‘Tell me, how was it?’. Secondly, clicking the ‘add activity’ button resulted after the improvement in the avatar asking whether the

participant would like to add the activity together or alone. When 'alone' is chosen the improved form is displayed. When 'together' is selected the avatar asks for each input separately.

3. *Make the 'activity type selection' field more prominent.*
Participants seem to miss to select the activity type at times. This may cause the preselected activity, school, to be wrongly assigned. The main reason for this behavior seems to be that participants started to fill in the form in different at different input elements. It was unclear whether a different lay-out, e.g. with bigger icons, would help. Therefore, the select activity type input field was left unchanged. Instead explaining and practicing the form in a structured fashion was added to the task list for the introduction session.

4. *Increase the user friendliness of the time field.*
Participants were struggling with selecting the time. Some participants, especially the younger ones, were not as practiced with the digital time notation. The primary reason of the struggle however was the dropdown input field where children could select the time. It took the participants relatively long, compared to the other input fields, the select the right time. Therefore, the dropdown input field was replaced with a horizontal slider. Users can now slide a pointer over a time axis to the right time.

5. *Make the purpose of the edit and delete buttons more clear.*
Participants did not immediately understand what the functionalities behind the pencil (edit) and cross (delete) buttons were. Adding a textual description to the icon buttons was considered. This however would affect the clean design. Instead, the design was left unchanged because the participants would get the meaning of those buttons after one try. Showing the functionality of the buttons was therefore added to the task list for the introduction session.

From the interview the following proposed interviews are identified:

6. *Make the 'add activity' form less cumbersome.*
Some participants complained about the 'add activity' form to be cumbersome. After the first multi-week user study the value of all the required input fields can be evaluated properly. The amount of input fields remained the same. Only the perception of how cumbersome it looks might be influenced. The changes implemented under point 2 have that same purpose. No further changes were made.

7. *Add the functionalities to automatically load glucose and insulin values.*
Participants indicated that they could store and view their glucose and insulin values on their pumps. They did not see the need to copy that to the diary. Automatically uploading the values into the diary would be a solution. In order to fulfill this request a partnership agreement was made with NetBasics. A software company that developed Gluconline, a digital diabetes diary for adults. They have a smartphone app that can retrieve the values from the pump and upload it to their server. MyPAL can request those values from their server with the right credentials. When a user has a Gluconline ID the values are retrieved upon login or after pushing the 'upload values' button. That button is only visible for users with a Gluconline ID.

8. *Add a tutorial where all the components and functionalities are introduced properly.*
The context of this proposal is that the participants were not given any instruction on how myPAL worked. They had to discover it for themselves. Most participants got the hang of it quickly. During

the main user study, the participants received a more thorough introduction. Creating a tutorial is a large exercise. Given the introduction were the functionalities are properly introduced a tutorial, by myPAL itself, was deemed unnecessary.

9. Add *more default activity types and add standard pictures that can be added to an activity.*
When opening the picture upload field, it would open the 'Pictures' folder on the local machine. During the pilot experiment this folder unintendedly contained some pictures of animals. The participants loved to browse these pictures and select them for their practice activities. This is the context of the request for standard pictures. Because myPAL wants to encourage children to take and upload their own pictures no standard pictures will be made available. Furthermore, there was not enough data available to create more default t activity types. That is why a functionality that allows users to add their own types was present in myPAL. The manually added activity types during the full user study will be used to identify more default activity types for a future iteration. In short, this proposed improvement was not implemented.

10. *Add all the required input more stepwise to prevent the form to be perceived as overwhelming at the first glance.*
This remark was mentioned in various forms by different participants. This is implemented by giving the option to complete the input fields one by one together with the avatar (see point 2.).

11. Add more color.
Designing an aesthetically appealing user interface is beyond the scope of this version of myPAL. It deserves serious attention [72]. This is left for a future iteration.

12. Add the functionality to add future activities and receive reminders for those activities.
This proposed functionality is a valuable addition to myPAL because it would make it a more mature a generally usable application. It would make myPAL a good candidate for replacing a standard calendar application. However, this improvement is left for a future iteration because it fell outside the scope of the set goals.

13. Add funny animations of the avatar.
A wide variety of avatar gesture were created for the full user study. While some are created to be experienced as comforting most have the purpose to be energetic, happy and hopefully fun.

14. Add the functionality to chat with the avatar.
Instead of using a form to collect all the necessary information a different tactic would be to retrieve the information from a conversation between the avatar and the child. This is a more natural interaction. Besides information retrieval the chat could also serve the purpose of increasing the feeling of relatedness. A simple question and answering conversation using the same principles as defined for responding to the added content would possibly make the avatar more relatable. However, such a functionality requires its own design cycle. The improvement is too big to implement for this project.

15. Add the functionality to give the avatar assignments such as "go stand on your head".
The participants indicated that they would like to see the avatar do funny gestures and movements, like dancing or standing on its head. Some even wanted the ability to command the avatar to this. This functionality falls outside of the scope of this iteration, because it both needs and deserves a dedicated design cycle. Questions for that design cycle are what gestures and movements are

appreciated and how will these movements be initiated e.g. on command or autonomously? A fundamental research question is what the effect will be of letting the children command the behavior of the avatar on its effectiveness to influence the children. This is left for future work.

6.2. User study

Once the myPAL software, as described in chapter 5, was finished and the improvements identified during the pilot were implemented a new group of children evaluated the myPAL system. Instead of a one-time interaction the user study had an evaluation period of three weeks. Another difference with the pilot was that the avatar could now respond to the added content.

The user study ran between 23 November and 17 December of 2015. Within that period the participants evaluated myPAL asynchronously (they started and finished at different dates) for at least three consecutive weeks. The study consisted of three parts: an intake session, a three-week period of myPAL use and an evaluation session. Both sessions took place at the diabetes care unit of the Gelderse Vallei hospital. MyPAL was used at home. In this section the used design, measures, materials, set-up and procedure is described. Furthermore, the results of the evaluation are presented and discussed in the last two sections.

6.2.1. Design

The goal of the user study was to observe how the participant interacted with myPAL, how their motivation to use myPAL changed over time and to measure their attitudes towards the avatar. The user study is designed to research the questions formulated in section 3.4.2.

All the participants interacted with the same version of myPAL under the same conditions. However, it is important to note that because every participant added different content the behavior of the avatar differed slightly per participant. Furthermore, the selection of appropriate avatar behavior was partly probabilistic which also caused a minor difference in behavior.

6.2.2. Measures

During the user study 'interactional data' of participants using myPAL were extracted from the log files. The interactional data consists of user behavior, avatar behavior and the added content. Furthermore, the motivation to use myPAL and the attitudes of the participants towards the avatar were collected using questionnaires and interviews. In this section both datasets are further introduced.

The measures discussed in this section are the concrete operationalization's of the causal model introduced in section 4.5.1. All the identified factors are represented in a factorial model displayed in the final subsection of this section.

6.2.2.1. Interactional data

When participants used myPAL every button press, page visit, et cetera is registered in a log file. The same holds for every behavior the avatar displayed. Furthermore, all the added content was stored. The behavior of both the participant and the avatar as well as the added content formed the set of interactional data. From the interactional data an *interaction timeline* was created for each participant. The interaction timeline actually consists of four timelines showing the behavior of the participant, one timeline displaying the behavior of the avatar and four timelines showing how the characteristics of the added content changed over time. Based on the interaction timeline

the performance measures, TOTAL ADDED CONTENT and CONSISTENCY, were calculated. The total amount of added content is calculated by adding all the counts for added activities, measurements, pictures and goals (see equation *(1*).

$$total\ added\ content = n_{activities} + n_{measurements} + n_{pictures} + n_{goals} \qquad (1)$$

The consistency is inversely proportional to the average amount of days between two consecutive items of content (see equation (2).

$$consistency = 1 / \frac{\sum_{j=1}^{n_{nonzero}} di_j - di_{(j-1)}}{n_{nonzero}} \qquad (2)$$

Where $n_{nonzero}$ represents the amount of days where content is added and di_j is the index of a day when content is added.

The behavior registered for each user can be divided in four categories: activities, measurements, pictures and goals. Those categories are related to the main functionalities of myPAL. For each category a separate timeline is constructed by stacking the counts for each selected behavior per day using a bar plot. The behaviors are listed in Appendix A 1.1 and the ones marked with an asterisk (*) were used in the timeline. The behaviors were registered by myPAL in the following fashion. Each time the user 'example' added a new activity a new line was added to the log file containing a timestamp, a label for the log item and the username: "2015-11-26 19:05:52.976, ADDED_ACTIVITY, example"

The behavior of the avatar was also logged for each participant. Each separate behavior as defined in chapter 5 was logged together with a timestamp and username after each execution (see Appendix A 1.2 for the complete list). Only the amount of behavior displayed by the avatar is taken into account in the avatar behavior timeline. A more extensive analysis of the behavior of the avatar is left for future work.

A different type of user behavior revolves around the content the participants added. In Appendix 1.3 the datasets containing the content are described. Instead of asking the question "how much did they add?" the question is "how rich is the added content?". This latter question is expressed by the following features extracted from the original datasets:

- Time per login
- Number of words per activity
- How personal is the activity?
  The following scale is used to establish how personal the description of the activity is:
    1. Not at all: a factual description of the activity.
    2. Minor: 1 + a minor indication of the thoughts/feelings of the participant.
    3. Moderate: 2 + information about friends and/or family.
    4. Major: 3 + major indication of thoughts/feelings about the activity and/or friends/family.
- Difficulty of goals
  To assign a goal participants needed to indicate either how many activities, measurements, etc. they wanted to add (daily goals) or how many consecutive days they would add an activity, measurement, etc. (total goals). The higher the target the more difficult the goal is.

The difficulty is calculated for daily goals and total goals separately. The range of the targets lies between 1 and 10 for daily goals and between 1 and 7 for total goals. The difficulty of the daily goals is calculated by adding the targets per day. The difficulty of total goals on each day is determined by adding all the current targets of each active total goal. The current target is equal to the original target minus the number of days.

6.2.2.2. Questionnaires and interviews

During the intake and the evaluations sessions the participant answered questions in two different fashions. The first type of questions was answered with the help of a 7-point Likert scale – *questionnaire* – and the second type of questions was part of a semi-structured interview – *interview*.

The 7-point Likert scale was used to be able to compare the results of this experiment to the results of the diabetes camp experiments where similar questions were asked using a 7-point Likert scale. Furthermore, during several child-robot experiments performed for the ALIZ-E project two important observations were made: 1) if children are presented with a list of questions some tend to give the same answer to all questions and 2) having to perform a physical act to answer a question seemed to stimulate the children to actively make a decision which benefits the truthfulness of the answers [32]. To take these observations into account each question was printed on a separate piece of paper, similar to Figure 31. The experiment leader presented the question by laying it in front of the participant and reading the question and scale indicators out loud. As a physical act the participant could choose one of the Playmobil® figurines, shown in Figure 32, and place it on one of the squares on the question sheet. The experiment leader presented the next question after writing down the given answer. The children had the tendency to point towards a specific number. When a figurine was placed ambiguously the experiment leader asked which number to write down.

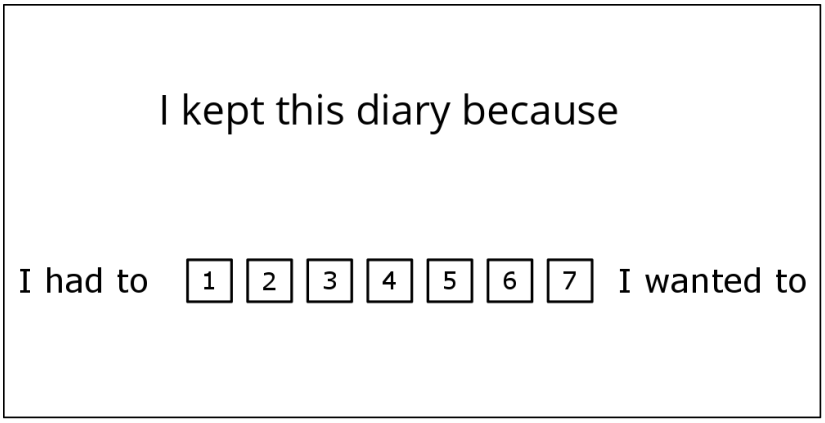


Figure 31: Example question of the questionnaire. The scoring was based on a bi-directional 7-point Likert scale.

Figure 32: Playmobil® figurines used to answer the questions in the questionnaire.

The goal of the avatar is to positively stimulate the intrinsic motivation of the children. To gain insight in the development of intrinsic motivation during the use of myPAL the following SDT-constructs were measured at the start (introduction session), halfway (via myPAL) and at the end (evaluation session):

- Feeling of AUTONOMY
- Feeling of COMPETENCE
- Feeling of RELATEDNESS
- General intrinsic MOTIVATION

To measure it at home using myPAL a button was placed half way on the calendar page that directed the participants to a digital questionnaire that had a similar appearance as it had in the introduction session. Instead of using Playmobil™ to select an answer the participants had to click the numbered box. Again the questions were presented one by one.

The TOTAL amount of MOTIVATION during the whole period was captured by the area under the motivation curve (AUMC; see equation (3)).

$$AUMC = \int_{pre}^{post} motivation(x)\, dx \qquad (3)$$

To make the total motivation more intuitive to interoperate it is normalized as shown in equation

(4.

$$Total\ motivation(x) = \frac{AUMC(x) - min_{AUMC}}{max_{AUMC} - min_{AUMC}}$$

$$where\ min_{AUMC} = 2\ and\ max_{AUMC} = 14 \qquad (4)$$

The scores for the feeling of AUTONOMY, COMPETENCE and RELATEDNESS are processed in the same way as for TOTAL MOTIVATION. The questions within each construct and the results of the reliability analyses using a Cronbach's Alpha score are discussed in Appendix A2.

Furthermore, the

- EXPERIENCE with computers and gaming

of the participants was collected during the introduction session. To be able to evaluate the avatar different questions falling under the following constructs were asked during the evaluation session:

- AVATAR BEHAVIOR CHECK
- PERCEIVED SOCIABILITY of the avatar
- APPRECIATION of the avatar behavior
- SELF-REPORTED IMPORTANCE of avatar

The questions are also listed in Appendix 2 together with the reliability analyses of the constructs. Besides motivation and attitude towards the avatar also a few demographical information that need to be included in the analyses, mostly to answer RQ4. In the questionnaire questions were included to collect:

- GENDER;
- AGE;
- DIABETES OFFSET in years and
- whether the children use a PUMP OR PEN to administer insulin.

Finally, during the introduction and evaluation session the children were interviewed in a semi-structured fashion. The core questions were prepared and asked, but the experiment leader had the freedom to ask follow-up questions to encourage the participants to elaborate their answers. The questions (fully listed in Appendix 2) collected information about:

- the EXPECTATIONS of (using) MYPAL (introduction);
- the EXPERIENCE WITH AVATARS AND ROBOTS (introduction);
- the EXPECTATIONS regarding the INTERACTION with the AVATAR (introduction);
- the EXPERIENCE of using MYPAL (evaluation) and
- suggestions for improving myPAL (evaluation).

Furthermore, during both the introduction and evaluation sessions the participants was asked to elaborate their answer for a subsection of the questionnaire questions (marked in Appendix 2).

6.2.2.3. Factorial model

In Figure 33 all the operationalized measures are displayed including all the expected effects. The expected effects are the following. All the avatar behaviors are observed by the participant (all the behavior checks pass) and contribute to the avatar being perceived as more social, its behavior being appreciated and rated as important by the participants.

Furthermore, a more positive attitude towards the avatar (perceived more as social, more appreciated behavior and rated as more important) lead to a higher support of the feeling of autonomy, competence, and relatedness as well as higher motivated participants that ultimately add more content consistently.

When the antecedents of motivation are supported the motivation itself is stimulated and participants perform better. The final expected result is that a better motivated participant leads to a better performing participant. All in all, resulting in a well-connected factorial model.

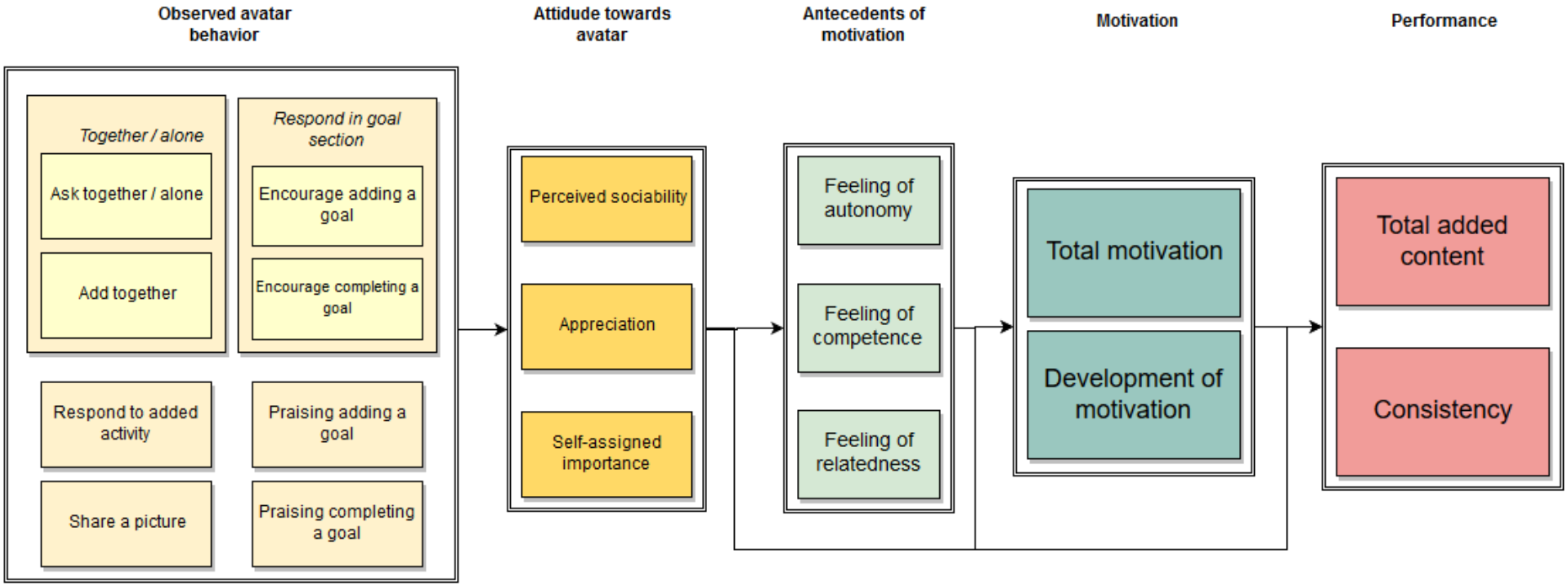


Figure 33: factorial model showing all operationalized measures and their expected relationships. The factors captured by a double lined box are related, but are analyzed separately. Connection to or from a double lined box indicate that all the factors involved are fully connected.

6.2.3. Participants

13 participants (age range: 7-12, median: 10; 4 girls, 9 boys), all patients at the Diabetes care unit of the Gelderse Vallei hospital, completed the experiment. There was one more potential participant who withdrew after the intake session. All participants were diagnosed with Diabetes Mellitus type I between 1 and 7 years ago. The participants were all under the supervision of one legal guardian. The participants had mixed levels of experience with regard to real robot encounters and diabetes diary use. All participants received a small gift and a certificate as a thank you for their contribution.

6.2.4. Materials and experimental set-up

MyPAL, as introduced in section 5, was hosted on a 4GB RAM 2CPU Windows Server 2012 R2 employing HTTP Strict Transport Security (HSTS). HSTS forced that the secure HTTPS (TLS 1.2) communication protocol always was used. This meant that all (medical) user data was encrypted at all times while being transferred.

The interview was recorded using the Smart Voice Recorder app (version 1.7.1), created by Andrew Kovalev, on a Samsung Galaxy S5 smartphone running Android 5.0.

As described in section 5 the avatar in myPAL is based on a red Nao robot. The participant could interact with the physical robot during the intake and evaluation session. The virtual and physical robot were both called Robin and were introduced as being the same entity. A red Nao (version 2.1.3) created by Aldebaran[3] was used in this experiment (see Figure 35). The Nao was controlled using a custom made Wizard of Oz (WoZ) application, similar to the one used in [5], running on a standard issue office laptop. The WoZ offered a list of selectable sentences that the robot could say. Furthermore, the robot could initiate and play a sorting game on touch table autonomously after activation by the experiment leader (also see Figure 35). The touch table was a touch enabled Samsung Series 7 700A7D 27-inch All-in-One Desktop PC (Intel Core i5-3470T 2.9GHz Processor, 6GB RAM, 1TB HDD, AMD Radeon HD7850M, Windows 8) The sorting game was only played during the intake session.

The room where both sessions took place consisted out of a desk where the participants (and their parents) and the experiment leader could be seated on opposite ends. This way the experiment leader could write down answers and other notes on the laptop and run the WoZ application without the participant seeing it. Furthermore, enough empty space was available to place the touch table with the robot and the participant on opposite sides. They were positioned in such a way that the participant was facing the robot and could not see the experiment leader using the WoZ application. A schematic overview of the room is shown in Figure 34.

[3] http://www.aldebaran.com/en

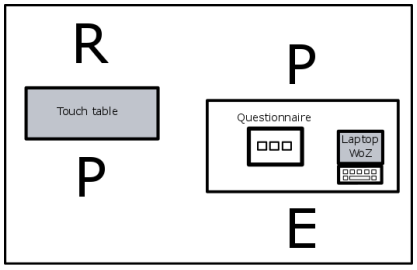


Figure 34: a schematic overview of the room where the experiment took place. On the right side the set-up of the questioning part of the intake and evaluation sessions is depicted. The participant (P) and the experiment leader (E) were seated across from each other. The questionnaire was presented on paper. On the left side the set-up of the interaction with the robot (R) is shown. The touch table was only present during the intake session. The participant was facing the robot and away from the experiment leader who was controlling the robot with the Wizard of Oz (WoZ) laptop.

6.2.5. Procedure

Intake session

The intake session had three parts: questions, introducing Robin the Robot and demonstrating myPAL. The experiment leader took the participant and a parent to a private room where the outline of the experiment and the formalities were discussed. After signing the consent forms the parent left the room. Meanwhile, the participant received instructions on how to answer the questions of the questionnaire. After two practice questions the participant answered all the 10 questions (see Appendix 2.1). The questionnaire was followed by a semi-structured interview. There was a fixed set of main questions that was asked in a specific order (see Appendix 2.1). The experiment leader however asked unprepared follow-up questions after each answer to get the participants to clarify, exemplify or deepen their original answer. The interviews were recorded on audio.

After all questions were done the participants were asked to fetch their parents while the experiment leader collected Robin the robot from the room next door. A towel was placed on the ground were the participant could sit. Robin was placed at eye level right in front of the participant. The touch table was placed between the participant and Robin (see Figure 35). The experiment leader explained that Robin was still sleeping and that he needed to be woken up. The experiment leader activated Robin using Wizard of Oz application. Using the WoZ set-up Robin had a conversation with the participant. At the end of the conversation Robin initiated the sorting game on the touch table. After 2 – 3 rounds Robin said that he needed to go and enjoyed its time with the child. Before carrying Robin out of the room the experiment leader offered the child an opportunity to get his/her picture taken together with Robin.

After Robin left the room the experiment leader loaded myPAL on a laptop. The child was given control over the laptop. The experiment leader introduced the main parts of the diary (calendar, goals and pictures). The child was given the assignment to add an activity to the calendar without much explanation. This way the children learned to use myPAL by their own exploration. The same principle was used to introduce the adding and clearing of goals.

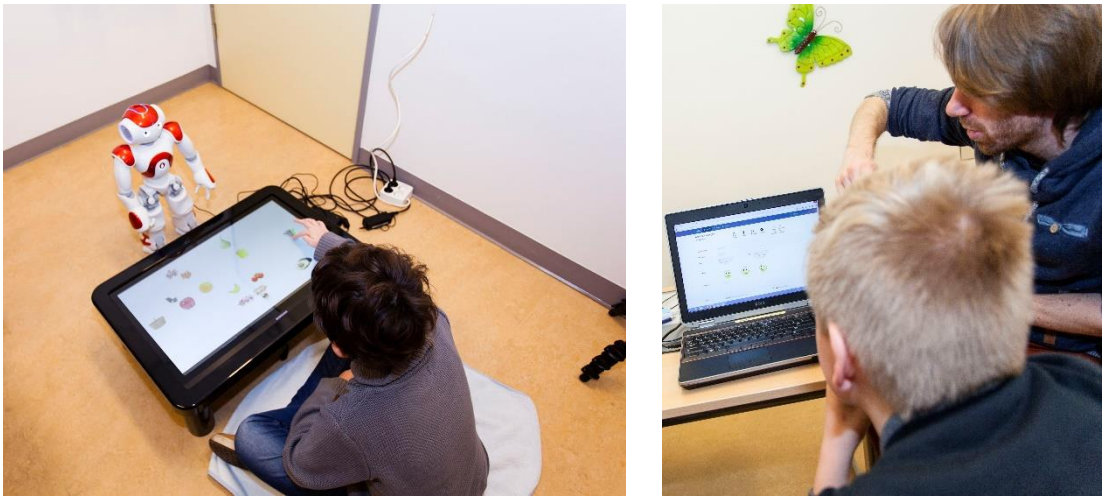


Figure 35: left: participant playing the sorting game together with Robin the Robot. Right: experiment leader introducting myPAL. Pictures by Rifca Peters ©

Using myPAL at home

Participants were instructed to access the myPAL web application using the latest version of Firefox on a laptop or computer. Although myPAL was fully-responsive (it adapted automatically to smaller screens custom to mobile devices such as tablets and phones) the speech audio file would not play on mobile devices.

During the intake session the parents were instructed to support their children but not force them. In practice this meant that the parents were allowed to ask whether the child already used the diary today but not make them so. Furthermore, the parents were allowed to help their children to log in or type a message only upon the request of the child. These instructions were given to make sure the participants only used myPAL intrinsically.

Most of the intake sessions took place before the official launch of the myPAL prototype. The participants were instructed to wait before they could login. The waiting time varied between three days and one day. All participants received an e-mail notifying them that they could login.

During the first few days three participant reported a minor bug. Two bugs could be fixed immediately and had little impact on the experience. The third bug could not be reproduced but a work around was provided minimalizing the influence it had on the experience. No big bugs or system crashes were reported.

Evaluation session

The evaluation session had three parts: questions, saying goodbye to Robin and thanking the participant. The procedure was similar to the intake session. The questioning part again consisted out of a questionnaire and an interview. The questionnaire contained 45 questions. The full questionnaire and interview questions can be found in Appendix 2.3.

The interaction with Robin differed on two points: 1) there was no touch table and thus no sorting game and 2) the content was about use of myPAL. For each child an activity that they added to myPAL was mentioned by the robot.

All participants received a certificate for their participation and a glow in the dark Hexbug™ nano to thank them.

6.2.6. Results

The results presented in this section are used to answer the research questions formulated in section 3.4.2. First the factorial model presented in Figure 33 will be evaluated by checking all the expected relationships. Secondly the participants are categorized based on their performance and motivation. By identifying distinguishing characteristics of those categories we further personalize the interaction children have with myPAL.

The strategy for analyzing the influence of all the factors represented in the factorial model is not a straightforward one. There are two issues that need to be addressed. The first issue revolves around the measurement scale of the factors based on the Likert questions, such as the behavior checks, user attitude, and motivation. In the literature there is a lot of discussion whether to consider the measurement scale as ordinal or as an interval. Technically speaking they are ordinal and they should be analyzed using non-parametric statistics. However, research shows that analysis of variance (ANOVA) is robust enough to treat Likert-based factor as an interval (continuous). This holds especially when the factor is constructed as a combination of several Likert questions [73].

The second issue is the low sample size. As a consequence, the power of statistical tests is low resulting in a higher risk for false negatives (falsely maintaining the null-hypothesis). The larger the influence of a factor the better the changes are of detecting it.

The aim of the statistical analyses is to validate the behaviors of the avatar as a beneficial addition rather than to construct a complete and truthful model of all the effects myPAL has on the child. The research is exploratory in nature rather than descriptive. The found effects, or the lack thereof, are considered as a starting point for future work rather than a full confirmation.

Given factor A and factor B we are interested in whether a high score in factor A leads to a high score in factor B (or the other way around). To facilitate as much power for the analysis the samples of factor A are divided in a LOW and HIGH (and sometimes MEDIUM) level creating a two/three-leveled independent between-subject factor. Factor B is treated as a continuous dependent variable.

One-way ANOVA's are used to test whether there are significant effects. Note that for a 2-leveled between-subject factor an ANOVA is equivalent to an independent t-test. SPSS is used to execute the tests. To account for multiple testing a Bonforonni correction has been applied on the p-values. These tests are only valid when the following assumptions are met:

1. The dependent variable is an interval or ratio variable. This assumption is met in all cases considering the discussion above.
2. The categorical variable should have two or more independent groups. This assumption is met.
3. The observations must be independent of each other, meaning that no participant can be in more than one group. This is mostly the case. When this is not the case it is because we are dealing with a repeated measures variable. A repeated measures ANOVA is then applied instead.

4. There should be no significant outliers. This is checked before every analysis. Any violations are reported. When there are no violations nothing will be reported to save space.
5. The dependent variable should approximately be normally distributed. This is tested with a Shapiro-Wilk (suitable for a low sample size) test before every analyses. ANOVA can handle small violations. Only large violations will be reported.
6. Homogeneity of variances is the last assumption that needs to be met. This is checked with a Levene's test. Violations are marked with an asterix (*) at the p-value. SPSS provides a correction when this assumption is violated. The corrected p-value is reported.

When there are multiple measures per participant, e.g. when evaluating the motivation timeline, a repeated measured one-way ANOVA is used to test the effect of the repeating measure. Instead of independent observations and homogeneity of variance, the variances of the differences between all combinations of related groups must be equal. This is called sphericity of the data. That is checked by Mauchly's test of sphericity. Once again, only violations are reported.

The factorial model is evaluated in steps. First the relationship of the observed avatar behavior with the other factors is investigated. This is followed by an analysis of the attitude of the participants towards the avatar. Subsequently, the antecedents of motivation and motivation itself are analyzed. The last step of the evaluation of the factorial model is check of effects of different demographical factors. Finally, the participants are categorized based on the performance and an attempt is done to identify predictors for those categories.

Observed avatar behavior

The avatar of myPAL employs several behavioral mechanisms to support the feeling of autonomy, competence and relatedness. How well these behaviors are observed by the participants are measured using behavior checks. The following behaviors are checked:

- TOGETHER OR ALONE: the avatar asking whether the child wants to add an activity together or alone and subsequently adding it together when the participants decides to do so.
- RESPOND TO ACTIVITY: the avatar responding to an added activity.
- RESPOND TO MEASUREMENT: the avatar did not respond to an added measurement. This is used as baseline.
- SHARE PICTURE: the avatar sharing a picture.
- RESPOND GOAL SECTION: when the child accesses the goal page the avatar engages in a conversation.
- RESPOND ADD GOAL: the avatar responding to an added goal.
- RESPOND MET GOAL: the avatar responding to a met goal.

In Figure 36 the mean scores for all the behavior checks are displayed together with their 95% confidence intervals. The question about the avatar responding to added measurements was included to represent a baseline for a “no / did not see” answer. Note that the avatar did not actually respond to the added measurements. It is furthermore noteworthy that children still scored the absence of the behavior with an average above 2. They seem to doubt whether they were not remembering it well or actually did not see the avatar respond to a measurement. This is not strange given all the other checks should be answered positively. Therefore, the other behavior checks are verified against this baseline instead of the null-baseline. To compare the behavior

checks with a baseline a repeated-measures ANOVA is performed. The reasoning is that if a check significantly differs from the baseline it is safe to say that the avatar behavior is generally noticed.

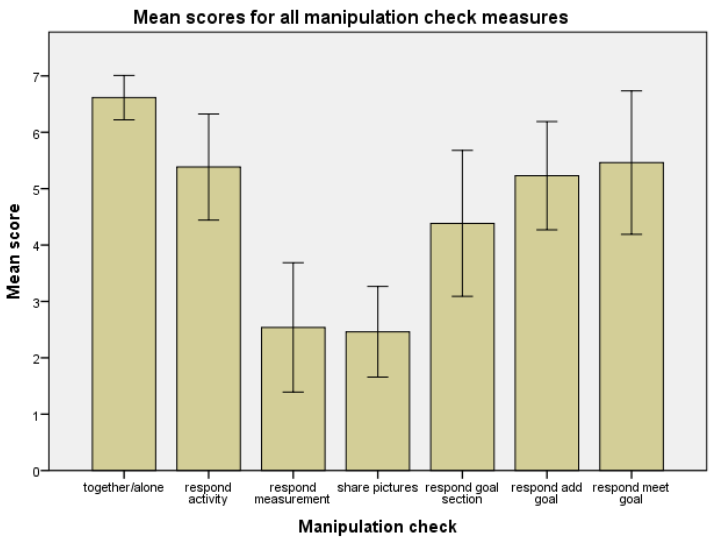


Figure 36: mean scores for all behavior checks with 95% confidence intervals.

All behaviors ($p < .027$) but the SHARE PICTURE and RESPOND GOAL SECTION ($p > .335$) differed significantly from the, RESPOND TO MEASUREMENT, baseline.

The scores of the behavior checks that significantly differ from the baseline are added together to capture the total conviction of the participants that they noticed the avatar behavior, called TOTAL OBSERVED BEHAVIOR (see Figure 37, left). To be able to check whether that conviction has an effect on the attitude towards the avatar of the participants the resulting score, TOTAL OBSERVED BEHAVIOR, is split in half. A natural separation can be identified in Figure 37 (left) as well. The two groups LOW and HIGH (see Figure 37, right) significantly differ from each other ($F(1,11) = 24.71,\ p = .0004,\ \eta_p^2 = .692, \pi = .995$).

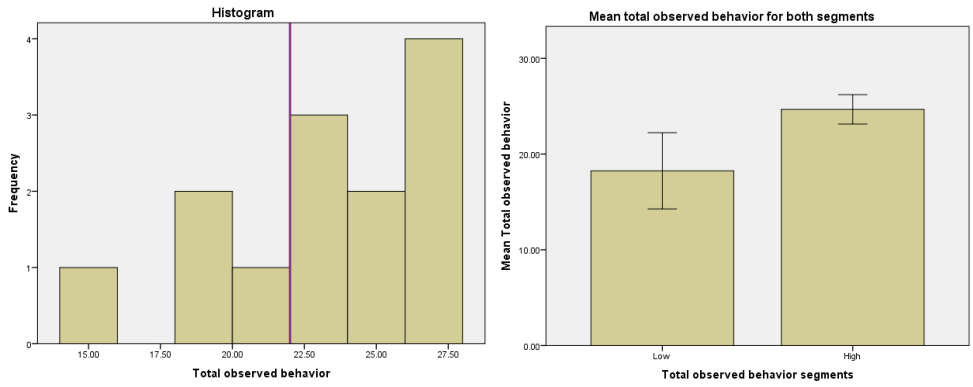


Figure 37: distribution of the total conviction the participants noticed the avatar behavior (left) and the means for both groups with 95% confidence intervals (right).

A main effect of TOTAL OBSERVED BEHAVIOR was found (see Figure 38) on SELF-ASSIGNED IMPORTANCE ($F(1,11) = 11.50,\ p = .006,\ \eta_p^2 = .511, \pi = .869$). Participants who were more convinced that they

saw the avatar rated the avatar as more important. A marginally significant main effect was found on the PERCEIVED SOCIABILITY ($F(1,11) = 4.4,\ p = .06,\ \pi = .482$). It appears that the more convinced participants also viewed the avatar as more social, but more research is needed to confirm this effect. No main effect was found on the APPRECIATION of the avatar ($F(1,11) = 2.03,\ p = .182,\ \pi = .256$).

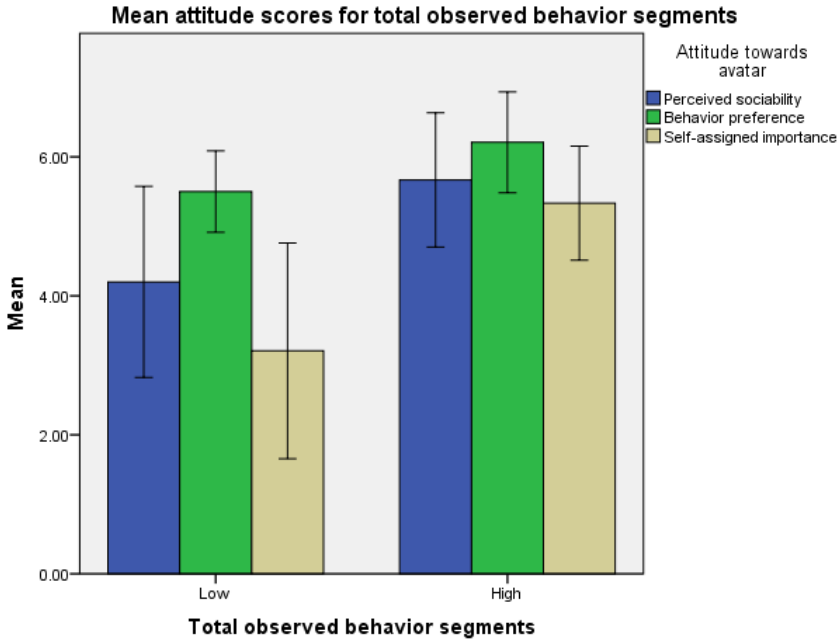


Figure 38: mean attitude scores for both total observed behavior groups with their 95% confidence intervals.

In Figure 39 the found results with respect to the factorial model are visualized. The greyed out factors are not included in the calculation of the TOTAL OBSERVED BEHAVIOR. A green line represents a positive effect. A solid line means that a significant effect was found. The lack of a line indicates that no effect was found and a dashed line indicates a marginally significant effect. These representations are used consistently for every factorial model figures that will follow.

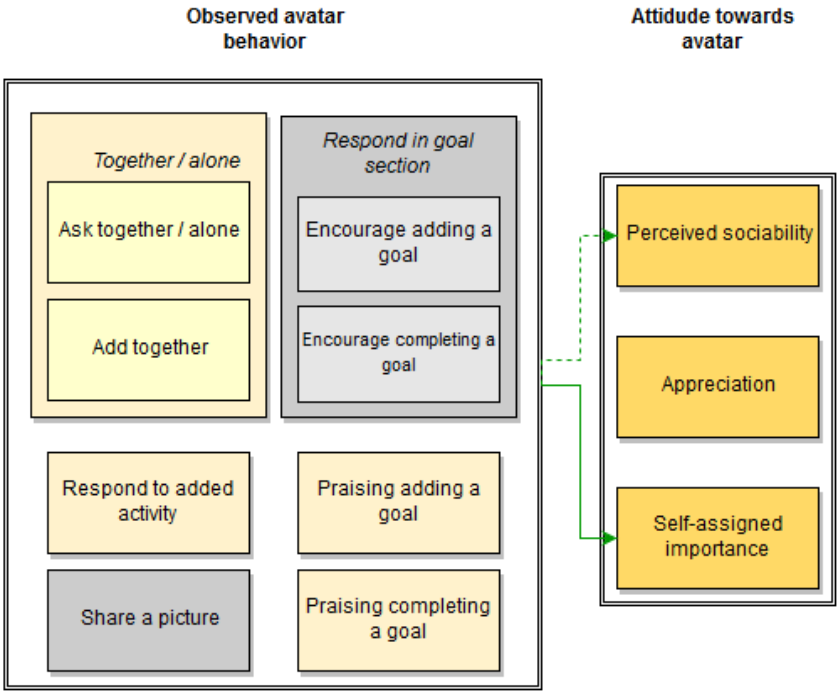


Figure 39: overview of which factors contribute to the TOTAL amount of OBSERVED BEHAVIOR. Furthermore, the figure represents the (marginally) significant main effects of the TOTAL OBSERVED BEHAVIOR on attitude towards the avatar. The effects that are greyed out do not contribute to the TOTAL OBSERVED BEHAVIOR. The solid green line represents a positive significant effect, the dotted green line a positive marginally significant effect. When no line is drawn no significant effect was found.

User attitudes towards the avatar

The attitude towards the avatar is expressed by three measures:

- PERCEIVED SOCIABILITY of the avatar by the participant;
- APPRECIATION of the avatar behavior by the participants and
- SELF-ASSIGNED IMPORTANCE of the avatar by the participant.

The distributions of all three of the measures are displayed in Figure 40. About half of the participants perceived the avatar as highly social while the other half perceived the avatar as not social at all to reasonably social (PERCEIVED SOCIABILITY). A simlar pattern is observed for the APPRECIATION of the behavior. Half of the participants highly prefered the behavior of the avatar while for the other half a variable preference for the avatar behavior was found. The SELF-RATED IMPORTANCE showed a more classical gaussian distrubtion were the ratings are more evenly spread.

The two groups (with LOW and HIGH scores) created for all three used attitude factors are displayed in Table 2. In Figure 41 the mean scores for the newly created groups are displayed. Furthermore, to verivy whether both groups difference significatly from each other a ANOVA with the group as independent variable and the score as dependend variable. The *p*-values are also displayed in Table 2. For all three attitude scores the two groups differed significantly.

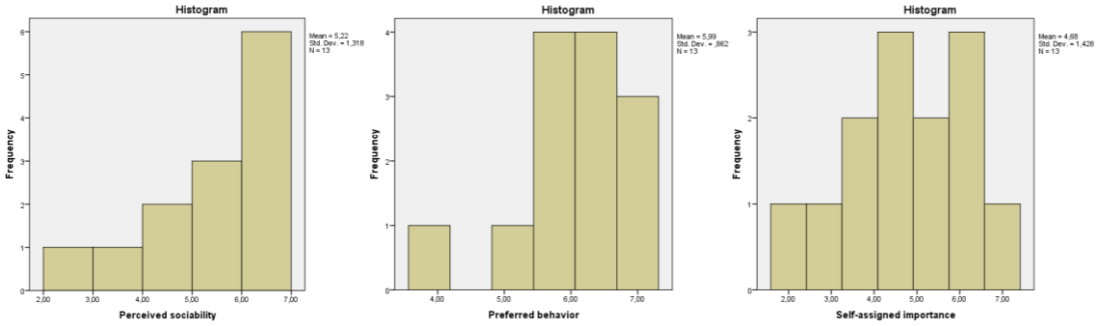


Figure 40: distribution of attitudes regarding the PERCEIVED SOCIABILITY, APPRECIATION *of the behavior and* SELF-ASSIGNED IMPORTANCE *of the avatar by the participants.*

Par	PS	A	SAI	Par	PS	A	SAI
1				*8*			
2				*9*			
3				*10*			
4				*11*			
5				*12*			
6				*13*			
				14			

Characteristics

	LOW			*HIGH*			
	N	*M*	*SD*	*N*	*M*	*SD*	*p*
PS	6	4.1	.44	7	6.2	.17	.001
PB	5	5.2	.79	8	6.5	.38	.002
SAI	7	3.6	.88	6	5.9	.75	.00004

Table 2: participants (par) are separated in two groups based on their PERCEIVED SOCIABILITY *of the avatar (PS),* APPRECIATION *of avatar behavior (A) and their* SELF-ASSIGNED IMPORTANCE *of the avatar (SAI). The groups with the lower rating are labeled yellow and the group with the higher ratings are labeled blue. An ANOVA with the group as independent variable and the score as dependent variable has been performed to determine whether the groups differ significantly.*

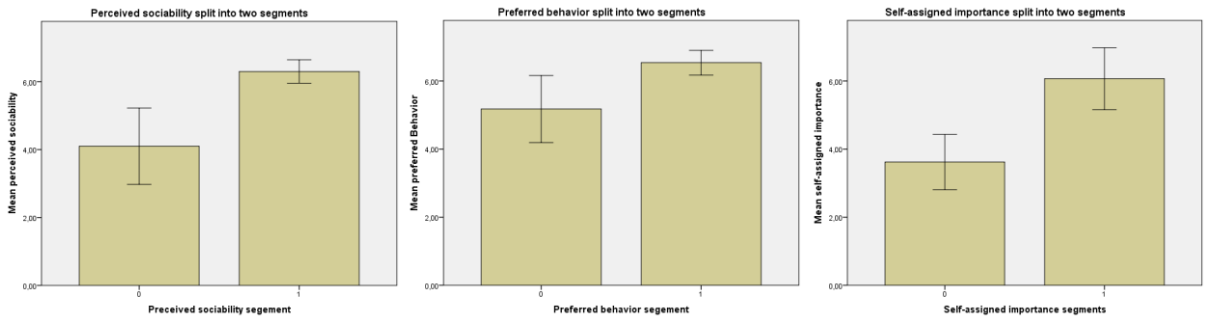


Figure 41: the mean scores of the PERCEIVED SOCIABILITY *(left),* APPRECIATION *of the behavior (middle) and the* SELF-ASSIGNED IMPORTANCE *(right) for both groups are displayed together with their 95% intervals.*

Performance

We will evaluate the performance of the participants for each of three attitude measures separately. In Figure 42 the amount of added content for all these groups are displayed.

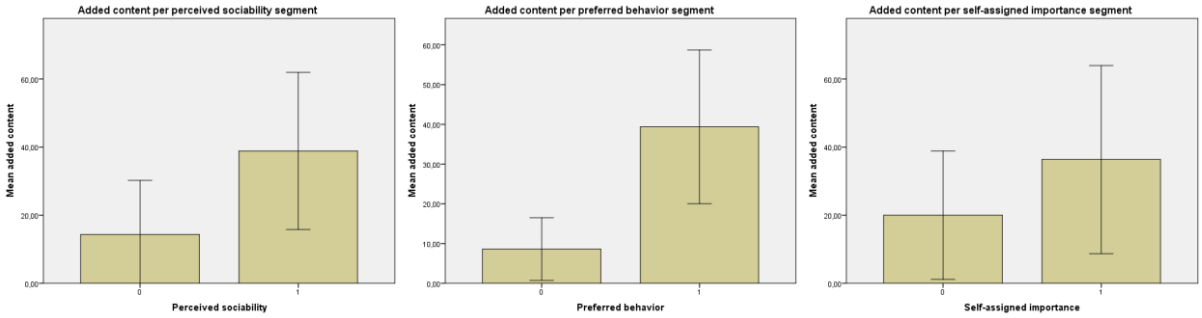


Figure 42: the total amount of ADDED CONTENT by the participants split in LOW and HIGH for the PERCEIVED SOCIABILITY (left), APPRECIATION (middle) and the SELF-ASSIGNED IMPORTANCE of the avatar (right) with their 95% confidence intervals.

A main effect of APPRECIATION on the ADDED CONTENT was found ($F(1,11) = 8.2,\ p = .007^{*},\ \eta_p^2 = .427\ and\ \pi = .741$). Participants who appreciated the behavior (HIGH) of the avatar added significantly more content than those participants who appreciated the avatar less - LOW ($\Delta_{content} = 31\ items$).

A marginally significant main effect of PERCEIVED SOCIABILITY on ADDED CONTENT was found ($F(1,11) = 4.4,\ p = .055^{*}\ and\ \pi = .48$). A trend is observed that participant who perceive the avatar to be more social add more content. The low power in relation with the low amount of participants and the high interpersonal variability is an indication that more nuance is required. More participants in a future study are needed to investigate this further.

No main effect is found of SELF-ASSIGNED IMPORTANCE on the amount of ADDED CONTENT ($F(1,11) = 1.6,\ p = .233\ and\ \pi = .211$).

In Figure 43 the CONSISTENCY per attitude group is displayed. The same procedure to check for main effects on the consistency have been performed.

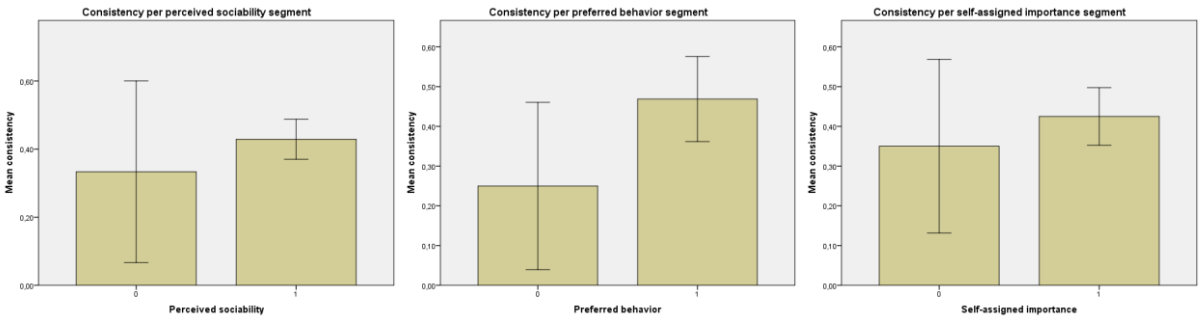


Figure 43: the average consistency for the participants with respectively a LOW and HIGH PERCEIVED SOCIABILITY (left), APPRECIATION (middle) and the SELF-ASSIGNED IMPORTANCE of the avatar (right) with their 95% confidence intervals.

A main effect of APPRECIATION on CONSISTENCY was found ($F(1,11) = 7.05,\ p = .022,\ \eta_p^2 = .391, \pi = .677$). Participants who appreciated the behavior of the avatar more (HIGH) also were more consistent ($\Delta_{consistency} = .219$). No other main effects were found ($All\ F's < .927,\ p's > .449^{*}\ and\ \pi's < .143$).

Motivation

The TOTAL MOTIVATION of the participants in all the attitude groups will also be compared (see Figure 44).

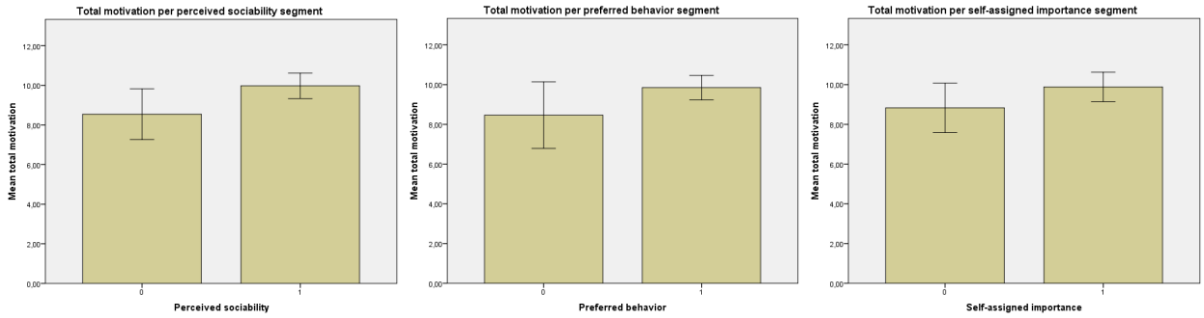


Figure 44: the total motivation for the participants in all the attitude groups (LOW VS HIGH, from left to right): PERCEIVED SOCIABILITY, APPRECIATION and SELF-ASSIGNED IMPORTANCE with their 95% confidence intervals.

A main effect of the PERCEIVED SOCIABILITY on TOTAL MOTIVATION is found ($F(1,11) = 7.03,\ p = .023,\ \eta_p^2 = .390, \pi = .676$). Participants who perceived the avatar as more social (HIGH) were more motivated to use myPAL ($\Delta_{motivation} = 1.43\ points$).

A main effect of the APPRECIATION of the behavior on TOTAL MOTIVATION is found ($F(1,11) = 5.89,\ p = .034,\ \eta_p^2 = .349, \pi = .600$). Participants who appreciated the behavior of the avatar more (HIGH) were also more motivated ($\Delta_{motivation} = 1.39\ points$).

No main effect of SELF-ASSIGNED IMPORTANCE on TOTAL MOTIVATION was found ($F(1,11) = 2.91,\ p = .116\ and\ \pi = .344$).

Antecedents of motivation

Besides investigating the relationship between the attitude towards the avatar and the general measure of motivation it is also interesting to investigate the relationship between attitude and the three SDT antecedents of motivation. In a similar fashion of how the total motivation is calculated the totals scores for the feeling of autonomy, competence and relatedness are also calculated (see Figure 45).

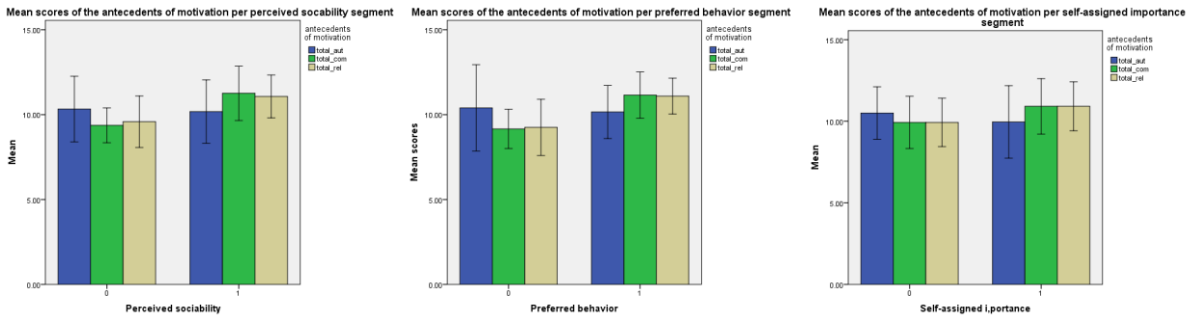


Figure 45: mean scores for the different antecedents of motivation per attitude group (LOW VS HIGH): PERCEIVED SOCIABILITY (left), APPRECIATION (middle) and SELF-ASSIGNED IMPORTANCE (right) with their 95% confidence intervals.

A main effect of PERCEIVED SOCIABILITY is found on the total FEELING OF COMPETENCE ($F(1,11) = 5.58,\ p = .038,\ \eta_p^2 = .337, \pi = .577$). The participants who rated the avatar as more social (HIGH) also indicated that they were supported in their competence. A marginally significant effect is found on the FEELING

OF RELATEDNESS ($F(1,11) = 3.63,\ p = .083\ and\ \pi = .413$). It seems to be the case that when participants viewed the avatar as more social (HIGH) they felt more related to it, but more research is required to investigate this observation. No main effect is found on the FEELING OF AUTONOMY ($F(1,11) = .021,\ p = .889\ and\ \pi = .052$).

A main effect of the APPRECIATION of the avatar behavior is found on both the FEELING OF COMPETENCE ($F(1,11) = 6.06,\ p = .032,\ \eta_p^2 = .355, \pi = .612$) as the FEELING OF RELATEDNESS ($F(1,11) = 6.26,\ p = .029,\ \eta_p^2 = .363, \pi = .626$). Participants who appreciated the behavior of the avatar more (HIGH) also felt more supported in their feelings of competence and relatedness. No main effect was found on the FEELING OF AUTONOMY ($F(1,11) = .049,\ p = .829\ and\ \pi = .055$).

No main effect of SELF-ASSIGNED IMPORTANCE was found on any of the antecedents ($All\ F's < 1.35,\ p's > .270\ and\ \pi's < .186$).

Factorial model

The results of the analysis of the attitudes of the participants towards the avatar with respect to the factorial model are displayed in Figure 46.

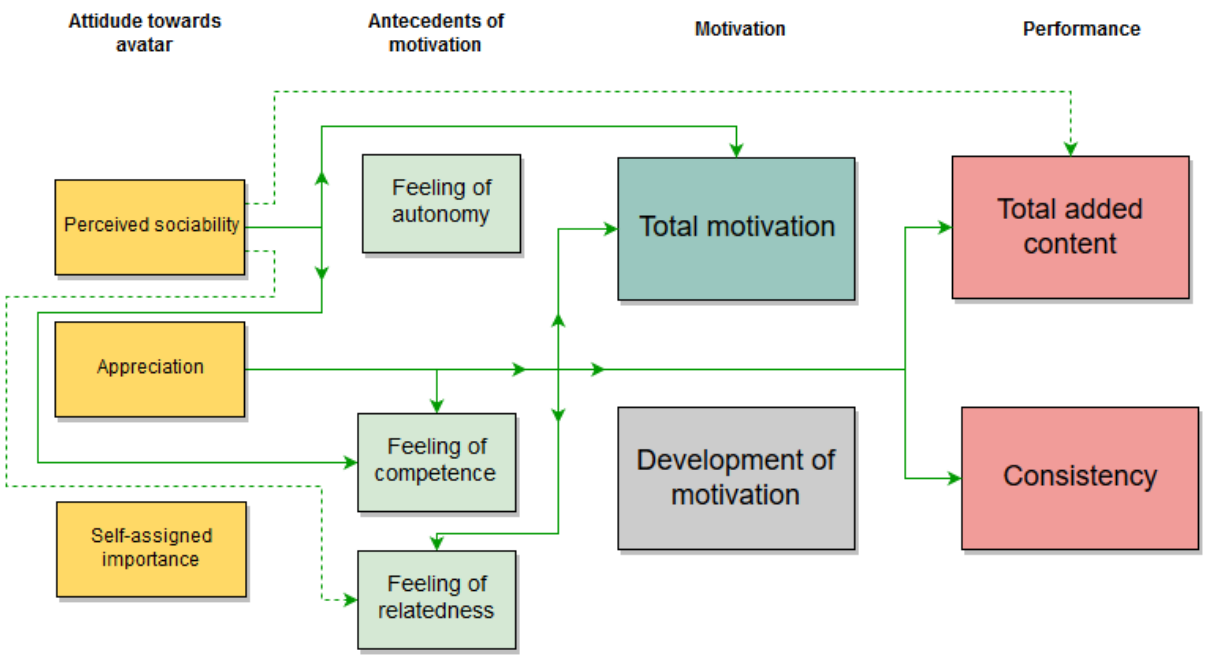


Figure 46: part of the factorial model with the effects of PERCEIVED SOCIABILITY, APPRECIATION, and SELF-ASSIGNED IMPORTANCE on the antecedents of motivation, TOTAL MOTIVATION, and performance inspected. DEVELOPMENT OF MOTIVATION is left out of the analysis (marked grey).

Antecedents of motivation

The distribution and the proposed groups of the antecedents of motivation (LOW vs HIGH) is displayed in Figure 47. In this paragraph we will investigate the influence of the feeling of AUTONOMY, COMPETENCE and RELATEDNESS on the TOTAL MOTIVATION and performance (ADDED CONTENT and CONSISTENCY).

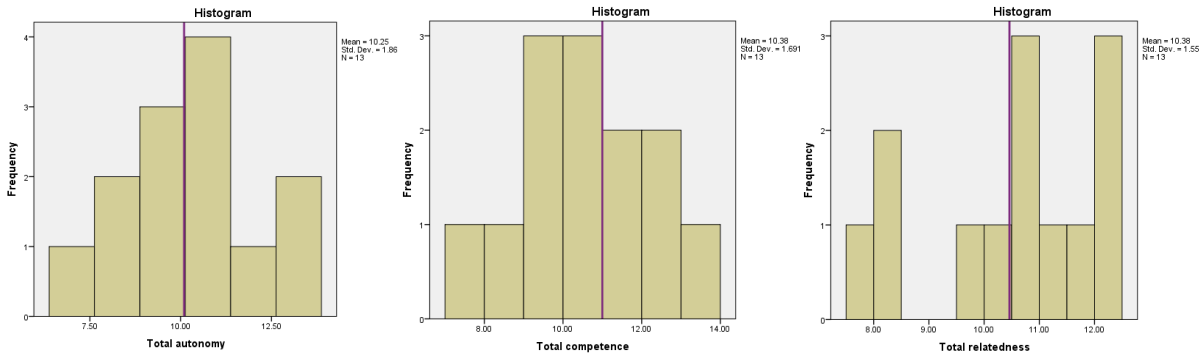


Figure 47: distribution of total scores of the feeling of AUTONOMY (right), COMPETENCE (middle) and RELATEDNESS (right)

In Figure 48 the mean motivation and the mean antecedent score is displayed for each of the three newly created groups. The difference of the LOW and HIGH scoring participangs significantly differ from each other for each antecedent ($F(1,11) > 12.85,\ p < .004, \eta_p^2 > .539, \pi > .903$).

A marginal main effect of AUTONOMY on THE TOTAL MOTIVATION is found ($F(1,11) = 4.17,\ p = .066,\ \pi = .462$). It is likely that participants who felt more autonomous were more motivated, but more research is required to confirm this.

A main effect of COMPETENCE on TOTAL MOTIVATION is found ($F(1,11) = 5.13,\ p = .045, \eta_p^2 = .318, \pi = .542$). Participants who felt more competent were more motivated.

A main effect of RELATEDNESS on TOTAL MOTIVATION is found ($F(1,11) = 7.43,\ p = .02, \eta_p^2 = .403, \pi = .699$). Participants who feld more related to the avatar felt more motivated.

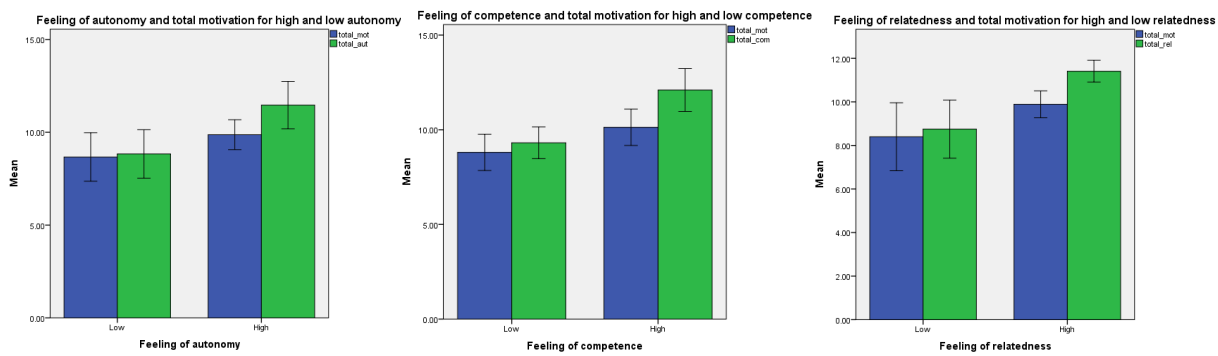


Figure 48: mean scores of TOTAL MOTIVATION (blue) and respectively TOTAL AUTONOMY (left), TOTAL COMPETENCE (middle) and TOTAL RELATEDNESS (right) in green for each motivation antecedent (LOW vs HIGH) with their 95% confidence intervals.

In Figure 49 the TOTAL amount of ADDED CONTENT for each antecedent of motivation is plotted. In Figure 50 the same figures are created but for the CONSISTENCY measure.

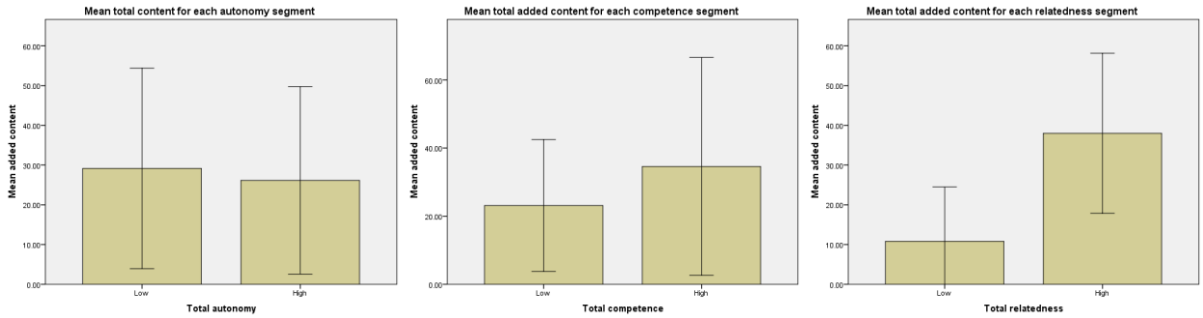


Figure 49: the amount of ADDED CONTENT for each antecedent of motivation groups (LOW vs HIGH): feeling of AUTONOMY (left), COMPETENCE (middle) and RELATEDNESS (right) with their 95% confidence intervals.

A main effect of the FEELING OF RELATEDNESS on the amount of ADDED CONTENT is found ($F(1,11) = 5.51$, $p = .019^{*}, \eta_p^2 = .334, \pi = .517$). Participants who felt more related to the avatar (HIGH) added significantly more content. No main effect of COMPETENCE and AUTONOMY is found ($All\ F's < .694,\ p's > .422\ and\ \pi's < .119$).

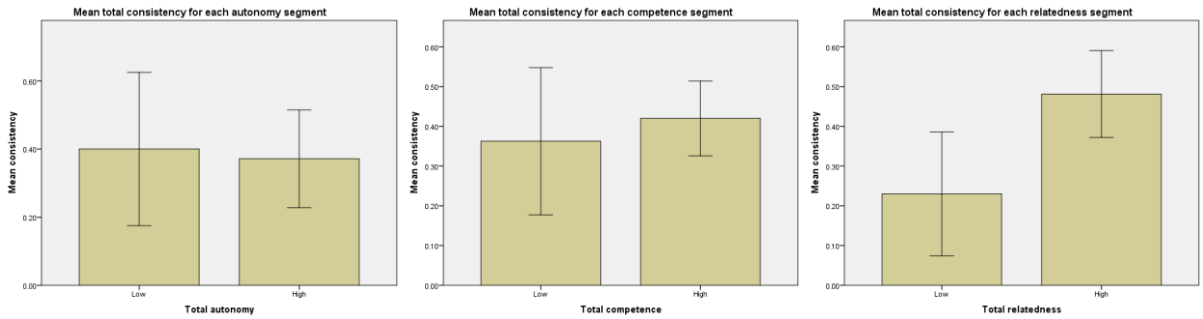


Figure 50: the levels of CONSISTENCY for each antecedent of motivation group: feeling of AUTONOMY (left), COMPETENCE (middle) and RELATEDNESS (right).

A main effect of the FEELING OF RELATEDNESS on CONSISTENCY was found l $F(1,11) = 11.7,\ p = .006, \eta_p^2 = .515, \pi = .875$). Participants who felt more related to the avatar (HIGH) added their content more consistently. No other main effects were found ($All\ F's < .305,\ p's > .517^{*}\ and\ \pi's < .080$).

The results of the analysis of the effects of the antecedents of motivation on motivation and performance with respect to the factorial model are displayed in Figure 51.

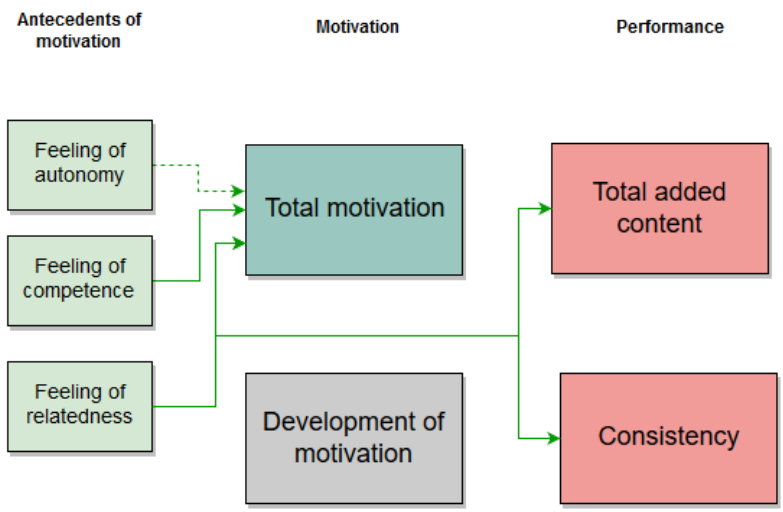


Figure 51: factorial model for the found effects of the antecedents of motivation on total motivation and performance. The development of motivation is left out of the analysis.

Motivation

One of the big subjects of this thesis is supporting motivation. When do we say we have succeeded in supporting motivation? In the first place the more motivation the better. But is that all there is to it? The way motivation develops over time is maybe even more important. If we consider the case of two hypothetical participants. The first is highly motivated at the start but loses his motivation along the way. The second is not motivated at all at the start but gets more and more motivated along the way. Even if the total level of motivation of the first participant is higher, I would argue that myPAL is a success for the second participant and a failure for the first one.

When looking more closely at motivation we will analyze both the development of motivation as the cumulative score. Motivation is measured at three moments in TIME. To investigate the development of motivation a repeated measures ANOVA is performed with TIME as within-subject factor and MOTIVATION score as dependent variable. No significant main effect of TIME on MOTIVATION ($F(2,11) = 2.288\ and\ p = 0.123, \pi = .419$; see also Figure 52) was found. The low power confirms that more participants are needed to draw generalizable conclusions. Categorizing the users may shed some more light on the different ways the motivation develops.

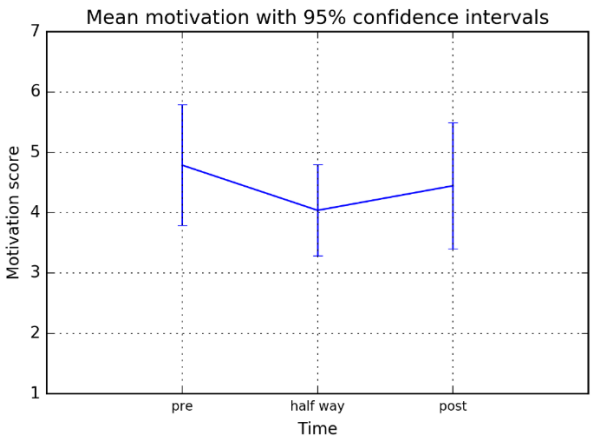


Figure 52: mean motivation of all participants combined together with the 95% intervals per measurement. No significant main effect of time on motivation was found.

To categorize participants based on their motivation development, timelines were made for all the participant. These motivation timelines, displayed in Appendix 4.2, show the scores on the questionnaires that measured the general motivation and the feeling of autonomy, competence and relatedness on all three measure moments in TIME (PRE, HALF WAY, POST). By visual inspection three main patterns in the motivational development were identified. One participant did not fit with either of the categories and will be treated separately (see Figure 53 bottom right). The first category of participants all share a general decline in motivation over time (see Figure 53 top left). The second category shows a V-shape in the development. Motivation drops between the first and second measurement and rises again between the second and last measurement (see Figure 53 top right). The third category displays no or little change in motivation over the measurements (see Figure 53 bottom left). In the first two groups an effect of TIME is to be expected and in the third group no effect of TIME on TOTAL MOTIVATION is expected.

Repeated measures ANOVA's, with a Bonforonni correction to account for multiple testing, have been performed within all three categories. In category 1 a main effect of TIME on MOTIVATION is found ($F(2,11) = 26.548$, $p = .001$, $\eta_p^2 = .898$, $\pi = .999$). The POST measurement was found to be significantly lower, difference of 1.85 points, than the PRE measurement ($p = .01$). No significant difference was found between the PRE and HALF WAY measurement and a marginally significant difference was found between the HALF WAY and POST measurement ($p = 0.83$).

In category 2 also a main effect of TIME on MOTIVATION is found ($F(2,11) = 11.104$, $p = .005$, $\eta_p^2 = .735$, $\pi = .940$). The HALF WAY measurement is significantly lower than the PRE (1.7 points) and POST (2.0 points) measurements ($p < .03$). No significant difference is found between the PRE and POST measurements ($p = 1.0$).

In category 3, as expected, no significant main effect of TIME was found ($F(2,11) = 0.70$, $p = .290$, $\pi = 0.94$). This category displays larger interpersonal differences compared to the other two categories. This is also reflected by the low power of this test. This is an indication that this category should be split into different, even more specific, categories. The low number of participants would make splitting the category a fruitless exercise from a statistical and generalizability point of view at this time. More research is needed to explore this issue further. Another observation is that although participant number 2 stood out during the visual inspection it would still fall within the boundaries of the current 95% interval of category 3. This would however further increase the variation in category three. Given all this information the three current categories remain unchanged.

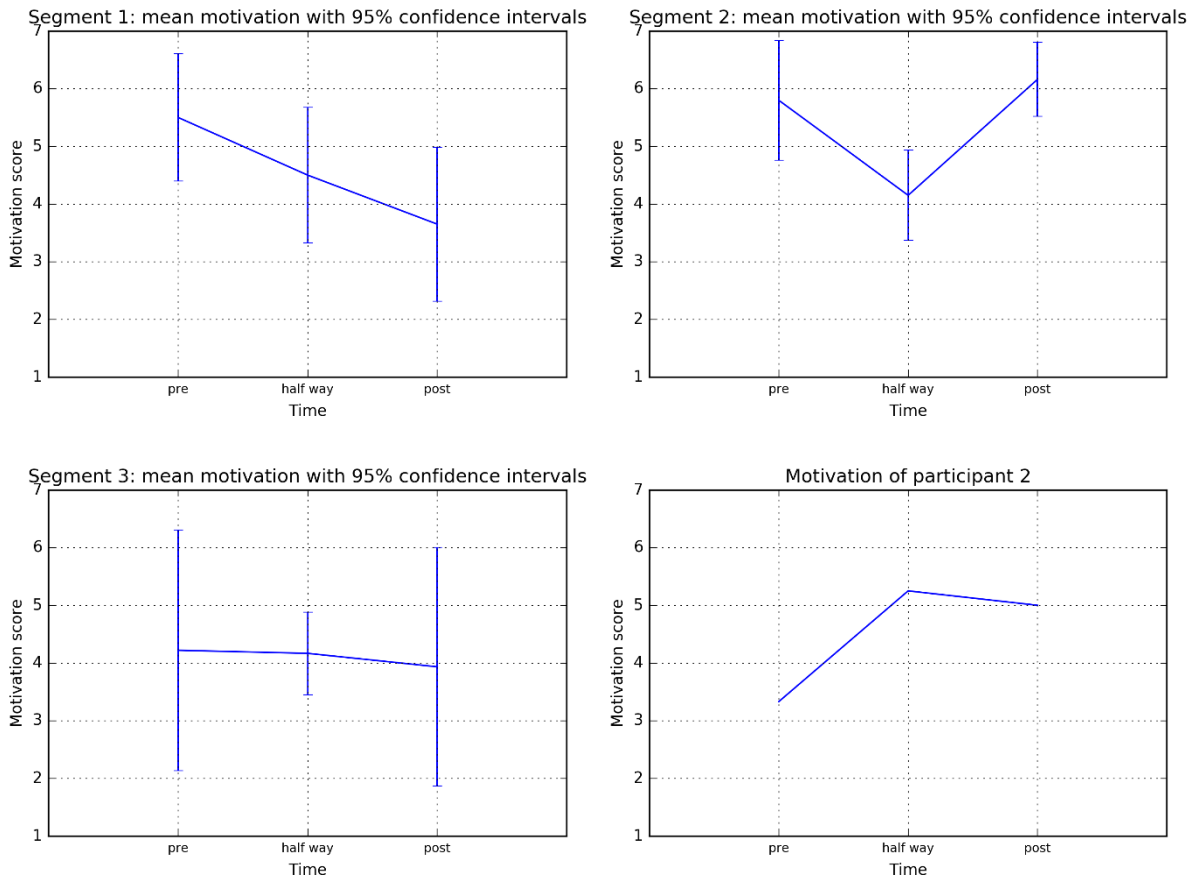


Figure 53: mean motivation with 95% intervals for each category. 1 top left, 2 top right and 3 bottom left. At the bottom right the motivational development for participant 2 is displayed who does not seem to fit in one of the categories.

Do these different temporal patterns result in a different amount of TOTAL MOTIVATION in the end? In Figure 54 the totals are displayed for each DEVELOPMENT OF MOTIVATION cateogy. An ANOVA has been performed to investigate whether there is a main effect of DEVELOPMENT OF MOTIVATION on TOTAL MOTIVATION. Results show only a marginally significant effect ($F(2,10) = 3.16,\ p = .091\ and\ \pi = .461$). Further inspection revealed that only between the second (V-shaped) and the third (no difference) category a significant different TOTAL OF MOTIVATION can be found ($p = .036$).

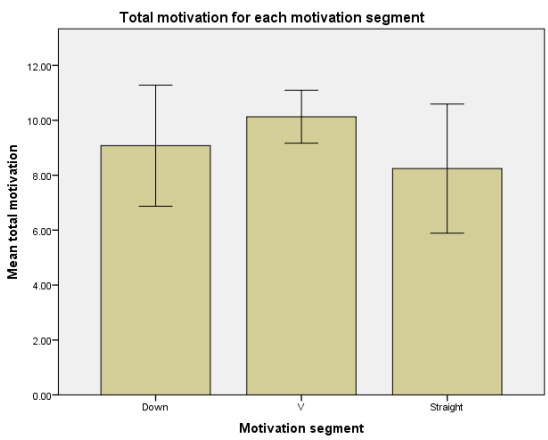


Figure 54: mean total motivation within each motivation category with their 95% confidence intervals.

Although the identified categories reveal interesting patterns of how differently motivation develops over time for the children, it is not responsible for, nor does it represent, the different total levels of motivation that the children displayed (see Figure 55, left). Therefore, a second set of categories needs to be created that represents that difference better. Looking at the histogram shown in Figure 55 a LOW, MEDIUM, HIGH categorization seems appropriate. With the first 4 participants receiving the LOW label, the middle 5 get the MEDIUM label and the last 4 receive the HIGH label. The scores for each category is also displayed in Figure 55 (right).

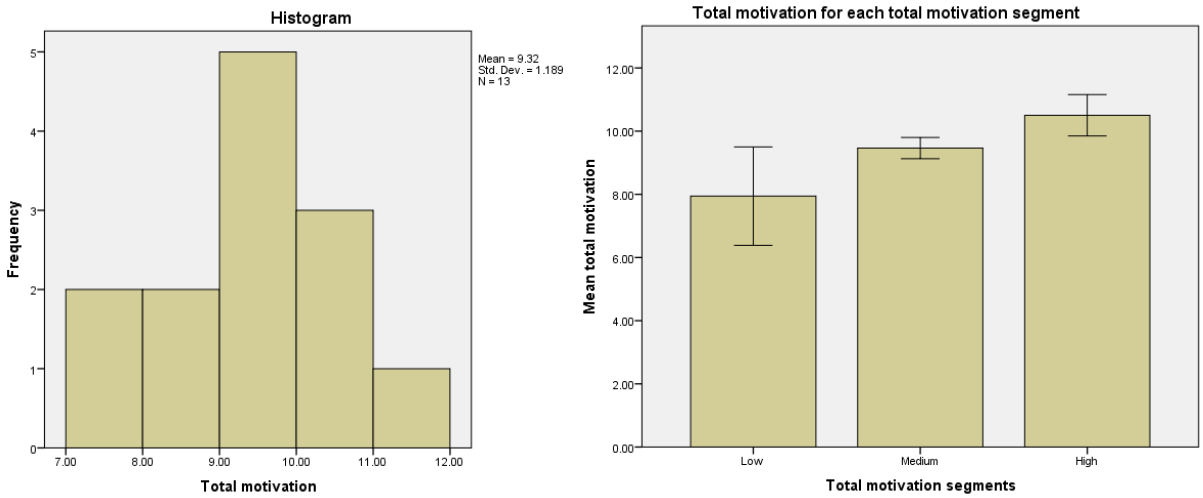


Figure 55: (left) histogram of TOTAL MOTIVATION showing the distribution of the final motivation scores and (right) the total motivation for the LOW, MEDIUM and HIGH categories with their 95% confidence intervals.

An ANOVA with the TOTAL MOTIVATION categories as between-subject factor and the TOTAL MOTIVATION as dependent variable showed that the created categories significantly differ from each other ($F(2,10) = 18.07,\ p = .0005, \eta_p^2 = .783, \pi = .997$). Further inspection revealed that the LOW significantly different from MEDIUM ($p = .012$) and HIGH ($p = .004$). Only a marginally significant difference was found between MEDIUM and HIGH ($p = .086$).

In Figure 56 the TOTAL MOTIVATION for each participant is plotted against their amount of ADDED CONTENT and level of MOTIVATION. A Pearson's correlation analyses revealed no significant correlation between the TOTAL MOTIVATION and ADDED CONTENT ($r = .287,\ p = .365$) and a marginal significant correlation between TOTAL MOTIVATION and CONSISTENCY ($r = .515,\ p = .087$).

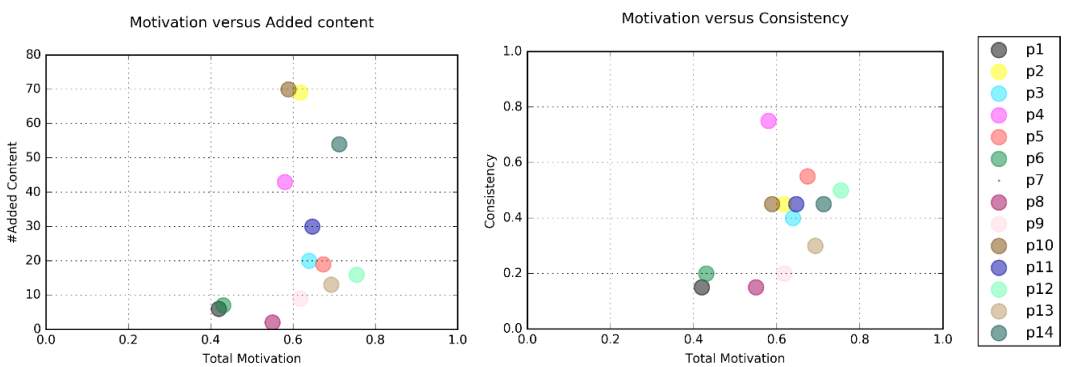


Figure 56: the total motivation for each participants is plotted against the total added content (left) and the consistency (right).

In Figure 57 the mean values of the performance measures are displayed for the MOTIVATION DEVELOPMENT. No main effect of MOTIVATION DEVELOPMENT was found on either the ADDED CONTENT as CONSISTENCY ($All\ F's < 1.07,\ p's > .384\ and\ \pi's < .182$).

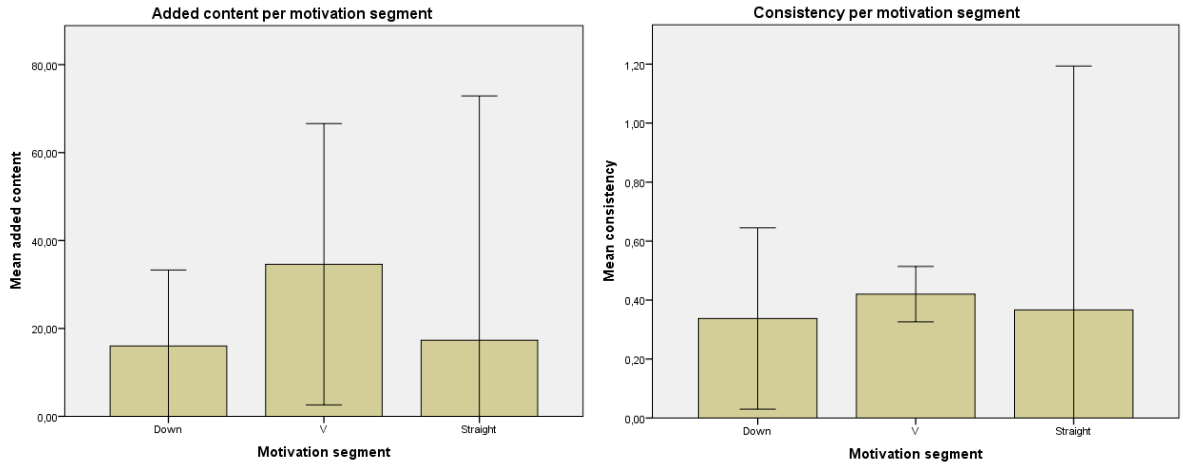


Figure 57: the mean values of the total amount of added content (left) and consistency (right) for each motivation development category with their 95% confidence intervals.

In Figure 58 the mean values of the performance measures are displayed for TOTAL MOTIVATION. Again no main effects of the TOTAL MOTIVATION on both performance measures were found ($All\ F's < 1.32,\ p's > .310\ and\ \pi's < .222$).

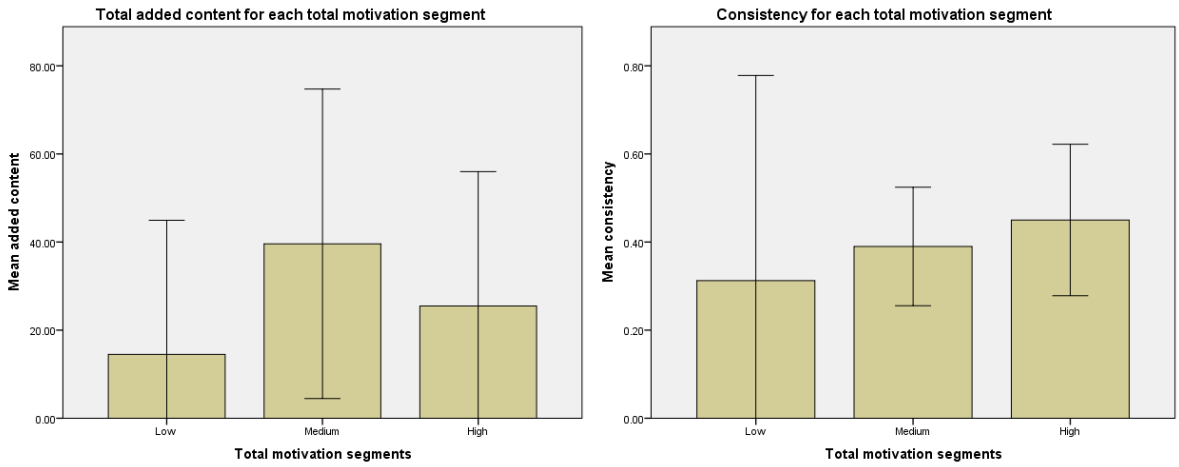


Figure 58: the mean values of the total amount of added content (left) and consistency (right) for each total motivation category with their 95% confidence intervals.

The results of the analysis of the (lack of) effects of motivation on performance is displayed in Figure 59. Note that there are no arrows because no effects have been found.

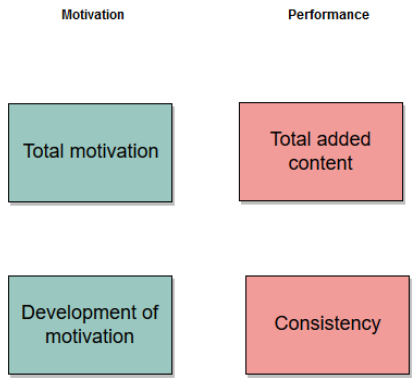


Figure 59: part of the factorial model showing (the lack of) effects of motivation on performance.

Demographics

Before we start looking different performance based categories and close the chapter on the factorial model it is good to check for effects of the different demographical factors such as GENDER, AGE, DIABETES OFFSET and whether the children use a PEN OR PUMP on the different factors.

An effect of GENDER on the FEELING OF RELATEDNESS is found ($F(1,11) = 7.69,\ p = .017, \eta_p^2 = .42, \pi = .729$). Boys felt more related to the avatar than girls ($\Delta_{relatedness} = 2\ points$). No other effects of GENDER are found $All\ F's < 3.12,\ p's > .105\ and\ \pi's < .365$.

No effect of AGE was found on any of the factors ($All\ F's < 1.14,\ p's > .309\ and\ \pi's < .164$). Furthermore, no effect of PEN OR PUMP was found either ($All\ F's < 2.95,\ p's > .114\ and\ \pi's < .348$).

An effect of DIABETES OFFSET on TOTAL OBSERVED BEHAVIOR ($F(1,11) = 5.30,\ p = .042, \eta_p^2 = .325, \pi = .555$) and SELF-ASSIGNED IMPORTANCE ($F(1,11) = 19.0,\ p = .001, \eta_p^2 = .634, \pi = .977$) is found. Children who were diagnosed with diabetes longer observed more of the avatar behavior ($\Delta_{TOB} = 2\ points$) and rated it as more important ($\Delta_{SAI} = 4\ points$). No other effects of DIABETES OFFSET are found ($All\ F's < 2.11,\ p's > .174\ and\ \pi's < .264$).

The end result of the complete factorial analysis is displayed by Figure 60.

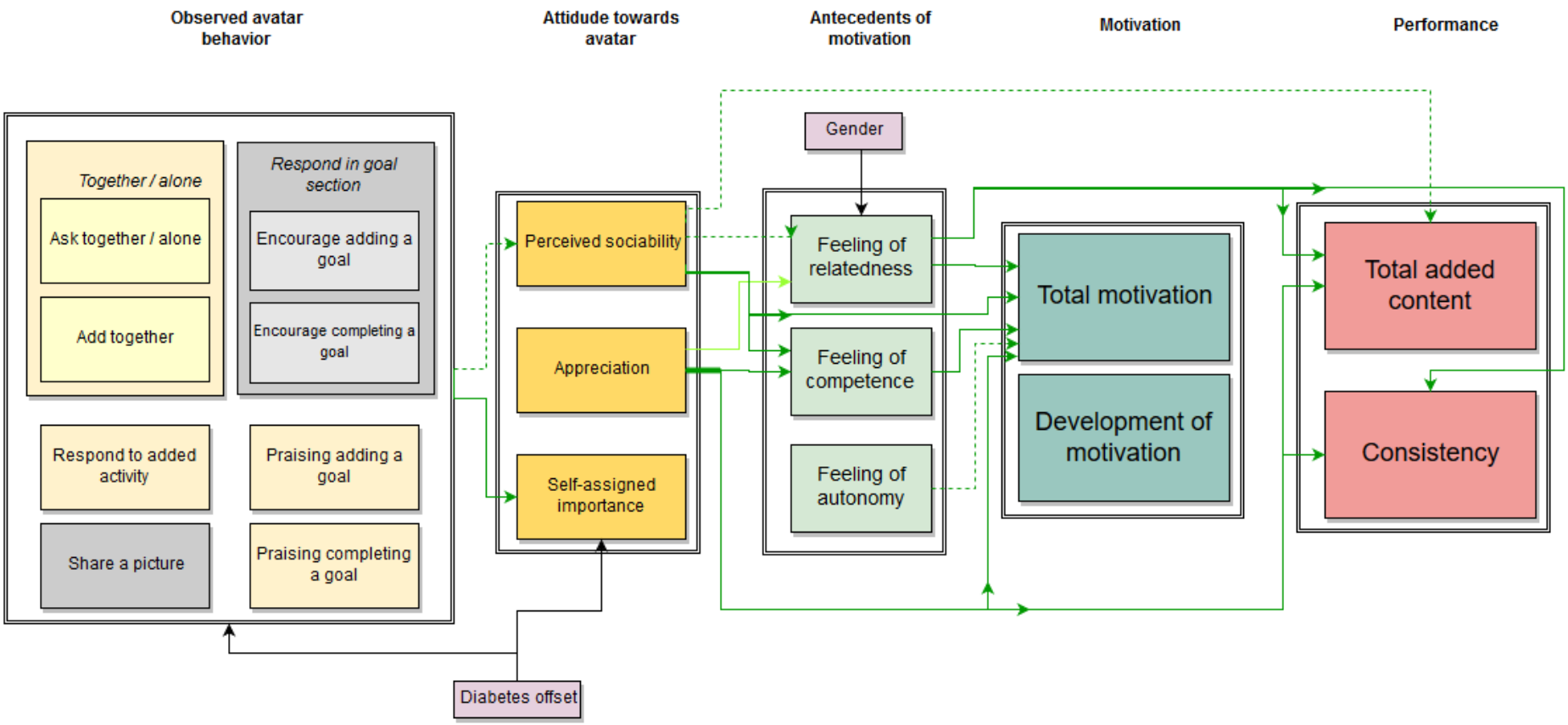


Figure 60: combined factorial model as a result of all the analysis on its parts. The non-greyed out behaviors are used to form the total observed avatar behavior. A solid arrow pointing from factor A to B indicates a significant main effect of factor A on B. A dotted arrow indicates a marginally significant effect. A green line indicates a positive influence (the higher the score on factor A the higher the score on factor B). There is furthermore an effect of gender on feeling of relatedness (boys feel more related than girls) and diabetes offset on self-assigned importance and observed avatar behavior (the longer someone has diabetes the more behavior that someone observes and the more important the avatar is rated).

Performance

To explore the development of the interaction of the participants an interaction timeline is created for each participant. The interaction timeline covers among other things the participant behavior towards content management over time. The interaction timelines for each participant are displayed in Appendix 4.1. From the timelines the total amount of ADDED CONTENT and the CONSISTENTLY of adding the content is calculated. The results are displayed in Figure 61. CONSISTENCY is plotted on the x-axis and the amount of ADDED CONTENT is plotted on the y-axis.

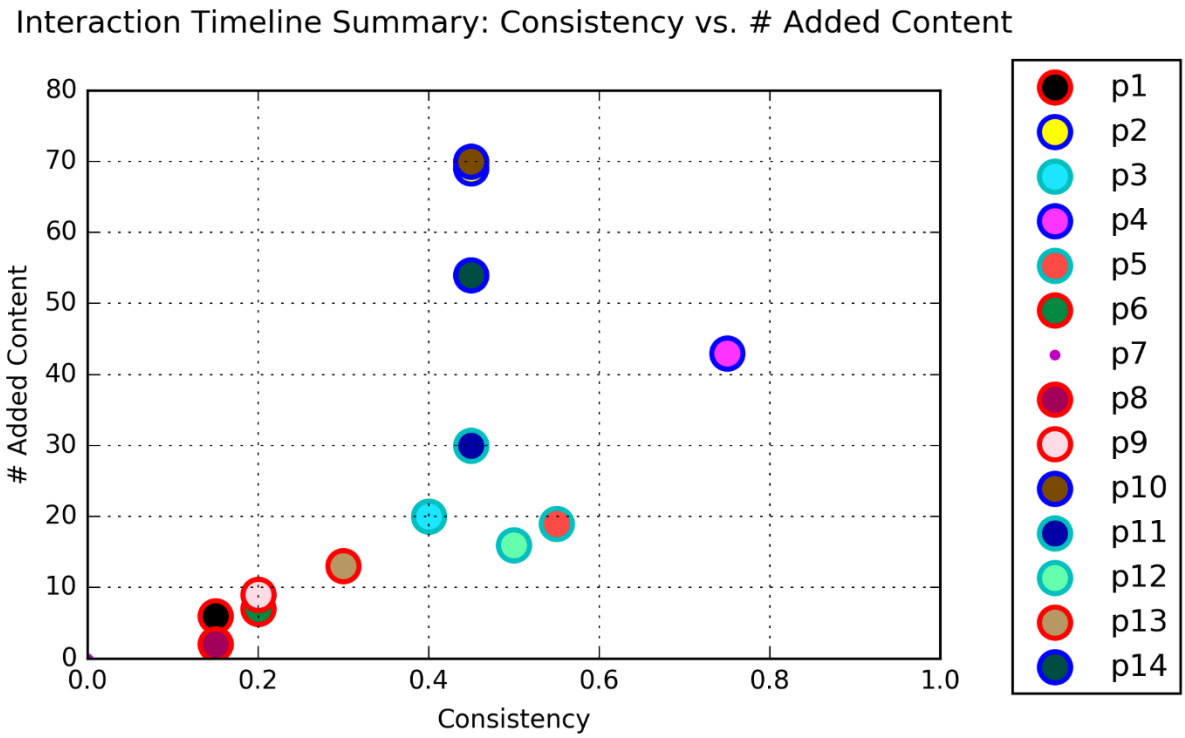


Figure 61: consistency plotted against the added content. Both values are derived from the interaction timelines. A low consistency implies longer gaps between logins. The added amount of content is the sum of the total activities, measurements, pictures and goals.

Categorization

To investigate which participants performed well and which participants did not perform as hoped we will categorize the children based on their performance. In order to capture both performance measures (ADDED CONTENT and CONSISTENCY) in the segmentation process both measures are clustered to create the categories. A standard *k*-means clustering algorithm implemented in the python Scikit-learn library [74] is used. Using the elbow method [75], were a balance is found between the number of clusters and the improvement in the score, it is determined that the number of clusters (*k*) can either be set to 3 or 4 (see Figure 62 respectively the green or red marker). It is not always obvious from the bat which elbow is the best choice [76]. However, another trade-off that plays a role can help determine a suitable amount of categories. Namely, the more categories the participants are divided in the lower the power of a statistical comparison becomes. Therefore, *k* is set to 3 which also has the benefit that participants are more equally spread over the categories. The participants were given a category label (LOW: red, MEDIUM: light blue and HIGH: dark blue) also displayed in Figure 61 by the border color of the marker.

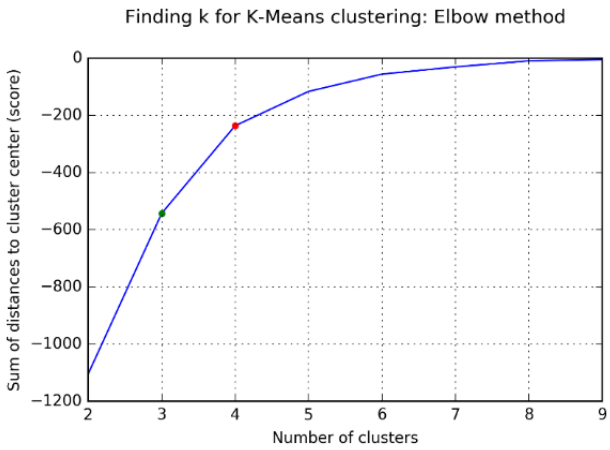


Figure 62: Elbow method for determining the k-value in k-Means clustering. The to be clustered data is the performances scores for the participants.

In Figure 63 the mean amount of ADDED CONTENT and the level of CONSISTENCY are shown per category together with their 95% intervals. For both the ADDED CONTENT and the CONSISTENCY a significant difference is found between the categories ($F(2,10) > 14.65, p's < .001, \pi > .986$). Further inspection revealed that for the amount of content the participants in category HIGH added significantly more content than the participants in category LOW (48 items, $p = .00004$) and MEDIUM (34 items, $p = .001$). A marginally significant difference was found between category LOW and MEDIUM (14 items, $p = .079$). The participants in category MEDIUM ($\Delta_{consistency} = .275,\ p = .007$) and 3 ($\Delta_{consistency} = .350,\ p = .003$) were significantly more consistent than the participants in category LOW. No significant difference between categories LOW and MEDIUM was found ($p = 1.0$).

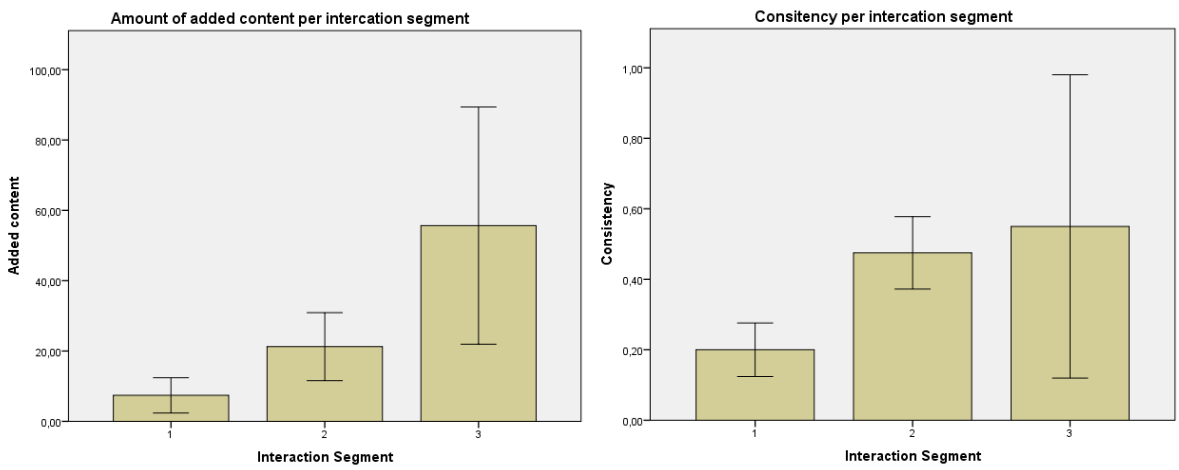


Figure 63: amount of added content (left) and the level consistency (right) for all three interaction categories with their 95% intervals.

Relationship ADDED CONTENT and CONSISTENCY

Before looking for factors that predict in which category a user will end up we will investigate the relationship between ADDED CONTENT and CONSISTENCY in more depth. A linear regression analysis shows that a more consistent user significantly adds more content ($F(1,11) = 5.72,\ p = .036$):

$$\text{ADDED CONTENT} = 79 \times \text{CONSISTENCY} - 2.72$$

When looking at Figure 61 this general pattern seems not to apply to all the categories. This is confirmed when the regression analysis is repeated per category instead. Only in the LOW category shows this linear relationship ($F(1,3) = 3.80,\ p = .032$). In the other categories no significant linear regression coefficient is found. There are more interpersonal differences between how much content participants add and how consistent they do that in the MEDIUM and HIGH categories.

Predictors

In Figure 64 the mean scores of all measured factors, who can possible play a role in predicting which category a participant will belong, are displayed for all three of the identified categories.

One-way ANOVA's with PERFORMANCE CATEGORY as between-subject variable and all the factors as dependent variables revealed that only the scores for DIABETES OFFSET significantly different among the categories (($F(2,10) = 4.18,\ p = .048, \eta_p^2 = .455, \pi = .594$). All other possible predictors had no significant different scores for the three PERFORMANCE CATEGORIES ($F(2,10) < 2.83, p's > .106, \pi < .432$).

Further analysis showed a marginally significant difference between category MEDIUM and HIGH ($p = .065$) for DIABETES OFFSET. No significant difference was found between LOW and MEDIUM (($p = .141$) and LOW and HIGH (($p = 1.0$). It appears to be the case that the children who have had diabetes longer are in the MEDIUM category and that the children who have had diabetes less long are either in category LOW or HIGH.

Due to the low sample size the chance for false negative is fairly high. This analysis should be performed after an experiment that is particularly directed at finding predictors for performing behavior that above all has more participants.

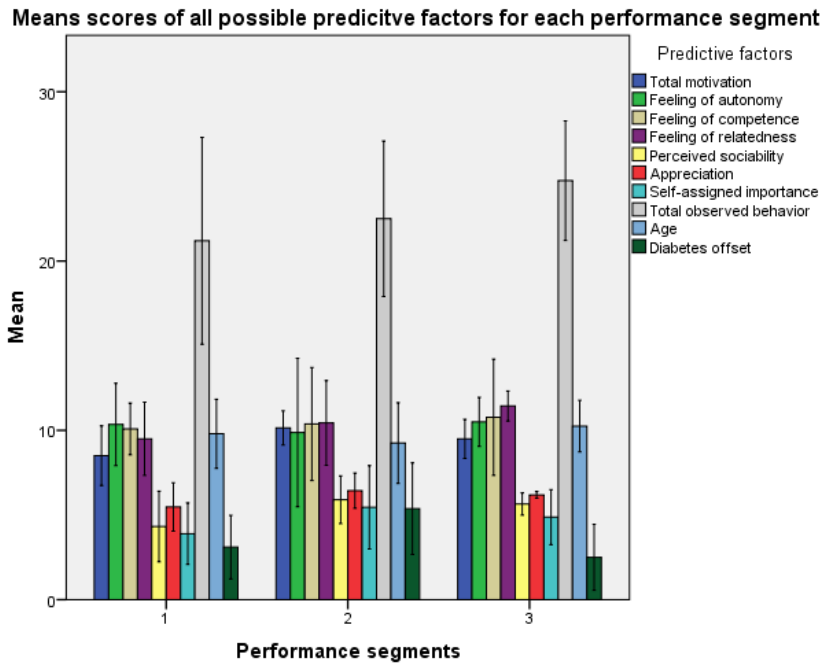


Figure 64: mean scores of all possible predictors within all performance categories with their 95% confidence intervals.

6.2.7. Discussion

The experimental set-up and the analysis of the results, presented in the previous section, are designed in such a way that they cover all the research questions formulated in section 3.4. In this section we will revisit the research question one by one and see if we can formulate an answer based on the results. Furthermore, we will discuss other interesting findings and issues that need further study.

The first question is:

RQ1. Which avatar behaviors are observed by the children?

We asked the participants to indicate how sure they were that they observed the following behaviors: TOGETHER OR ALONE, RESPOND TO ACTIVITY, SHARE PICTURE, RESPOND GOAL SECTION, RESPOND ADD GOAL and RESPOND MET GOAL. We also asked the same question for RESPOND TO MEASUREMENT while in fact the avatar did not respond after a measurement was added to myPAL. RESPOND TO MEASUREMENT was used as a baseline in the analysis with the following reasoning. If a behavior check significantly differs (positively) from the baseline the participant was significantly more sure that it observed the behavior. The behaviors that we consider to be significantly observed are TOGETHER OR ALONE, RESPOND TO ACTIVITY, RESPOND ADD GOAL and RESPOND MET GOAL. SHARE PICTURE and RESPOND GOAL SECTION did not pass the check.

Participants indicated in the interviews that they did not or hardly got to see a picture of the avatar. This is confirmed by the avatar behavior logs. Due to a technical flaw the pictures were not shown to the participants until after the half way mark of the experiment. The opportunity for the children to observe a shared picture was far lower compared to the other behaviors. This fits with the results and is the most likely explanation. Note that sharing a picture was not the only behavior facilitating mutual self-disclosure. Also responding to added activities contained that feature.

The avatar behavior logs however did reveal that the avatar responded most of the time when the participants accessed the goal section. A possible explanation of why this check failed is the relatively abstract formulation of the question. The other checks specifically ask for a concrete behavior, e.g. behavior after goal was added. The RESPOND GOAL SECTION behavior asks to verify whether accessing the goal page resulted in avatar behavior. Two different behaviors might occur, namely motiving the children to complete an active goal or to add a new one. The children might have found the question confusing which explains the lower score.

All in all, most avatar behaviors seem to be picked up by the participants.

The second question is:

RQ2. What is the effect of the amount of observed behavior on attitude towards the avatar?

The more the participants observed the avatar behavior (TOTAL OBSERVED BEHAVIOR) the more important they rated the avatar (SELF-ASSIGNED IMPORTANCE). It also appeared that observing the avatar behavior more resulted in a higher rating of its sociability (PERCEIVED SOCIABILITY), only this was marginally significant. Observing the behavior more did not make the participants appreciate it more (APPRECIATION). An important design conclusion is that seeing the avatar more does not mean that participants appreciate it more, only that they think it is more important.

In human-robot interaction it not uncommon that participant explicitly claim that the role of the avatar or robot is substantial while their implicit behavior or ratings say otherwise [77]. After answering the next question, we can verify whether this discrepancy between explicit and implicit assessment is the case by checking if participants who claim the avatar is more important (SELF-ASSIGNED IMPORTANCE) also perform better (add more content consistently).

The third question:

RQ3. What is the effect of attitude towards the avatar on the antecedents of motivation, motivation itself and performance?

The attitude of the participant towards the avatar is captured by three different measures:

- PERCEIVED SOCIABILITY: how social is the avatar perceived in the eye of the participant.
- APPRECIATION: how much is the behavior of the avatar appreciated by the participant.
- SELF-ASSIGNED IMPORTANCE: how important do the participants rate the avatar.

This research questions contains three separate questions. The first is whether these three attitudes influence how autonomous, competent and related to the avatar the children feel. The results show that none of these attitudes have an effect on how autonomous the participants feel. This is not unexpected because most of the autonomy supporting strategies are in the diary itself. The few autonomy supporting strategies employed by the avatar are clearly not sufficient for that purpose.

However, participants who appreciated the behavior of the avatar feel more competent and more related to the avatar. Participants who perceived the avatar as more social also felt more competent and appeared to feel more related to the avatar although that last result was marginally significant. The avatar behaviors implemented to support competence seem to their job although they are more effective if the avatar is perceived as more social. This connection is not strange because the strategy for autonomy support is based on giving out praise by the avatar. Praise is also a feature that is considered as social behavior [27].

The second and third sub questions revolve around the influence of user attitude on motivation and performance respectively. The results show that participants who perceive the avatar as more social are more motivated to use myPAL and ultimately add more content. The motivation supporting behaviors are mostly social in nature. Let's assume that when participants are more influenced by the social behavior of the avatar they perceive it as more social. Given the result that when an avatar is perceived as more social participants are more motivated and add more content we can indirectly conclude that the combination of all behaviors achieve their goal in motiving the participants to add more content.

Furthermore, the results also show that participants who appreciate the behavior of the avatar are more motivated, add more content and do that more consistently. Whether or not someone appreciates the avatar behavior is a key factor for the effectiveness of the avatar. In order to further personalize the behavior of the avatar more research must be done to identify which behaviors are appreciated by the different users.

Finally, no effect of self-assigned importance was found on any of the other factors. Coming back to the previous question, we found that observing the avatar more (TOTAL OBSERVED BEHAVIOR) makes

it appear more important (SELF-ASSIGNED IMPORTANCE). This intuition the participants have is not correct if we look at motivation support and performance, because when participants rate the avatar as important they do not necessarily are better motivated or add more content.

The fourth question:

RQ4. What is the effect of the antecedents of motivation on motivation and performance?

As one would expect, given the large body of evidence for the self-determination theory and the use of validated questionnaires, a higher feeling of autonomy (marginally significant), competence and relatedness result in more motivated participants.

Furthermore, only a higher feeling of relatedness results in participants adding more content and do that more consistently. This stresses on the one hand the importance of relatedness support, which is a design recommendation for a future iteration of myPAL. On the other hand, it indicates that either the feeling of autonomy and competence are not supported enough to play a role or play no role at all. This is a question that is left for future research.

The fifth question:

RQ5. What is the effect of motivation on performance?

No effects of motivation are found on performance. Given the result that several factors both stimulate motivation and the performance it is unlikely that motivation does not influence the performance at all but that its effect is too weak with the current set-up. The low power of the analysis indicates a high risk for false negatives. An explanation of why the effect of motivation is too weak is that it is a combination of the feeling of autonomy, competence and relatedness. Since the feeling of autonomy and competence also show no effect of performance the overall strength of the motivation as a factor is too low.

The sixth question:

RQ6. What is the effect of user demographics on the amount of observed behavior, user attitude, antecedents of motivation, motivation, and performance?

The following demographics are checked: GENDER, AGE, DIABETES OFFSET and whether the children use a PEN OR PUMP. Results show that boys feel more related to the avatar than the girls. Since the feeling of relatedness seems to be the main conduit between avatar behavior and more content being added consistently, it stresses the need for a more gender specialized interaction between the avatar and the participant. Van der drift et al. (2014) found the same result [5].

Furthermore, children who were diagnosed with diabetes longer observed more of the avatar behavior and rated it as more important. A possible explanation is that children who are dealing with diabetes have more experience with these kinds of experiments and know what to pay attention to. Information about the previous experiences is collected during the interviews and supplied by the parents. This is left out of the scope of this thesis but is available for further analysis.

The seventh question:

RQ7. What categories of children can be identified based on the performance?

With k-means clustering three categories have been identified. In the first category participants are captured that are not very consistent and do not add much content. It is therefore labeled as the LOW category. In the second and third category the relatively MEDIUM and HIGH performing participants are captured.

Results show that as a general rule the more consistent the participants are the more content they add. However, adding a lot of content does not necessarily mean that participant do that equally consistent. Starting from the MEDIUM category and especially in the HIGH performing category the linear relationship between added content and consistency cannot be significantly established.

The eight question:

RQ8. What categories of children can be identified based on the development of motivation?

Three categories were identified based on the development of motivation. In the first category the participants share a general decline in motivation over time (see Figure 53 top left). The second category shows a V-shape in the development. Motivation drops between the first and second measurement and rises again between the second and last measurement (see Figure 53 top right). The third category displays no or little change in motivation over the measurements (see Figure 53 bottom left).

The 'declining' category was expected because there are always participants for who motivation support is not effective for whatever reason. The 'stable' category was also expected. These are the participants for who the motivation support probably is effective. Motivation is generally high at the start because participants volunteered to participate with the experiment. Keeping that stable is considered a success. However, the V shaped category was not expected. A plausible alternative explanation is that the different condition of the HALF WAY questionnaire, which was answered digitally from home instead of with the figurines at the hospital, causes the difference in score rather than a drop in motivation.

The development of motivation did not have any effect on the performance. This finding makes the alternative explanation more plausible because a drop in motivation should result in other effects such a lower amount of added content for example. This cannot be completely confirmed because the total level of motivation also did not have a significant effect on performance. A 'fake' motivation drop for some participants might play a role in the lack of an effect of total motivation as well. If the total motivation of some participants was indeed higher than measured it could distort the results. This possibility should be taken into account for the following iteration. A half way measurement at the hospital would be a safer choice for measuring the development of motivation.

The ninth and final question:

RQ9. What are the characteristics of the identified categories in terms of performance, motivation, demographics, and attitude towards the avatar?

In Figure 64 the means of all the relevant factors within each performance category is displayed. Only the years someone has diabetes (DIABETES OFFSET) was found to be different in the categories. It appears that participants who have had diabetes longer are more present in the MEDIUM category compared to the LOW and HIGH categories.

The goal of this analysis was to find which factors are able to predict in which category a participant would end up. The aim for that exercise is to ultimately find out what the ideal conditions are for high performing participants. This question can also be answered by looking at factors who have a positive influence on ADDED CONTENT and/or CONSISTENCY.

7. Conclusion

In this thesis I set out with two major goals. The first was to identify specifications for a digital diabetes diary with a responsive avatar that is able to motivate children to add more content consistently. Especially focusing on the behavior of the avatar. The second aim was to evaluate the effectiveness of myPAL and its ability to motivate children to add more content consistently. Again focusing on the avatar behaviors.

In Table 3 an overview is given of all the implemented avatar behavior, how it is measured and whether it is successfully picked up by the children. This is an extract of Table 1. Table 1 shows all the myPAL functionalities (that provides an answer to design questions DQ1 and DQ2). Here, with Table 3, we focus on the avatar behavior in order to evaluate that more thoroughly.

Functions avatar	Interaction design pattern	Human Factor Requirement	Measure	Observed
Ask to add activity together or alone	Question with answer options	AUT6	TOGETHER OR ALONE	Yes
Add activity together	Ask activity item	AUT5	-	Not directly measured
Respond to added activity	- Mood matching - Question with answer options - Open question - Remark - Self-disclosure - Praise	REL2 REL1 REL1 REL1 REL3 COM1	RESPOND TO ACTIVITY SHARE PICTURE	Yes No, due to technical flaw
Encourage adding a goal	Encourage	AUT3	RESPOND GOAL SECTION	Not sure
Encourage completing active goal	Encourage	AUT3	RESPOND GOAL SECTION	Not sure
Praise adding a goal	Praise	COM1	RESPOND ADD GOAL	Yes
Praise completing a goal	Praise	COM2, AUT4	RESPOND MET GOAL	Yes

Table 3: overview of the functions of the avatar, how it is measured and whether it was observed by the participants.

Most behaviors are picked up apart from the avatar sharing a pictures and the avatar encouraging the adding or completing a goal. The first behavior was not picked up do to a technical flaw. The most plausible explanation for the last behavior is that the children did not properly understood

the evaluation question rather than not observing the behavior. These behaviors are left out of further discussion.

Results show that the implemented avatar behaviors that are picked up by the participant, positively influence the attitude towards the avatar. The more positive the attitude towards the avatar is, the more motivated the children are to keep their diary and the better they do it. This brings us to the first main conclusion:

***The avatar is a useful addition to diabetes care:* the results show that the avatar is able motivate children to add more content consistently.**

There are however two constraints that modulate that success. The first starts with an interesting result that states that children who observe the avatar more think the avatar is more important. This however need not be the case. Because children who think the avatar is more important are not more motivated nor do they add more content.

Appreciating the behavior of the avatar on the other hand is the key factor for being more motivated to add more content consistently. Another important result in this context is that observing the avatar more does not make participants appreciate it more. The second main conclusion to draw from this is:

***Quality over quantity*: the avatar behavior must be appreciated by the children in order to be effective. Simply showing the avatar more does not increase the appreciation. The behavior must match the children's preferences.**

The avatar is able to effectively support the feeling of competence and relatedness under two conditions. Once again, the behavior must be appreciated by the participant. Secondly, it must be perceived as social. The avatar behaviors that support the feeling of competence and relatedness are social in nature such as praise, mood matching and self-disclosure. The behaviors implemented for autonomy support, that were observed, had a less social character, namely the avatar presenting the choice how to add the content. The lack of the social nature of the autonomy support behavior is a plausible explanation of why the avatar was not able to effectively support that. The other way around is also likely, the more social the behavior is the more effective it is to support motivation and ultimately lead to better performing children.

That social behavior is an important factor for the children also became clear after the pilot study. The most requested features for myPAL where to ability to chat with the avatar and the ability to let the avatar make funny gestures and movements.

Furthermore, a strong correlation ($p = .86$) between appreciation and perceived sociability suggests a relation between the two. Although no causal direction can be inferred from a correlation result, it is clear that avatar sociability and appreciation go hand in hand. This brings us to the third and last main conclusion:

***Avatar sociability is key*: the more social the behavior of the avatar is, the more it is appreciated by the child and the more motivated the children are to add more content consistently.**

The results also show that the current implantation of the avatar does not provide the right behavior for every child i.e. not every child appreciated the behavior or perceives it as social. More research is needed to explore more specifically which behaviors are preferred by the different children. In this research the behaviors are evaluated as a whole instead of separately. They however provide a starting point for a more separate analysis. Steps to find predictors for identifying low, medium and high performance of the participants were unsuccessful. A larger study is required to proceed further.

When looking at the whole process, from design to evaluation, we can conclude that this research contributes to both diabetes care as child-robot interaction. This research contributes to diabetes care by providing the specifications (and a prototype) of a diabetes diary with a responsive avatar that is able to motivate children to add more content consistently. The content the children add to the diary is useful for the children to gain insight in their diabetes and the relation between food, mood and blood glucose values. This insight allows them to anticipate better how to react to (upcoming) hyper's or hypo's. In other words, it improves their self-management skills. Research shows that a better self-management plan leads to a better quality of life for children with diabetes [1].

This research also contributes to field of child-robot interaction by combining two relatively novel approaches. The first is the autonomous behavior of the avatar. By far most of the similar studies used a wizard of Oz set-up to control the behavior of the avatar. The biggest danger is that the apparent social and cognitive capabilities of the avatar cannot be contributed to the avatar itself but belong to the one controlling the avatar. A difference might exist between, for example, the perceived sociability of the avatar when the avatar operates autonomously or under a 'wizard'. This would allow for a behavior to be deemed effective while in fact the effectiveness of the behavior is not attributed to the behavior itself but to how it was applied by the wizard. This danger does not exist in this research because the avatar operates autonomously. All the drawn conclusions can only be attributed to the avatar itself and not a human operator.

The second approach is a longer term study. Instead of looking at a one-shot interaction, which is common practice in (c)HRI, this study evaluates the effectiveness of the interaction over a period of three weeks. This is still too short to evaluate the effectiveness of myPAL and the avatar in a realistic setting, because that would require months or even years of interaction. It however is more likely to surpass the gimmick effect of the avatar, making the results more robust, than a one-time or very short user study.

This research brings the sustainable application of autonomous avatars a step closer to being realized, especially in the field of diabetes care and cHRI. If this research teaches us anything it is that a social avatar goes hand in hand with a high performing child, where quality of the avatar behavior rather than the quantity determines the success.

8. Acknowledgements

This thesis would not have been as good as it is now without a lot of people. First of all I would like to thank my direct supervisors for their guidance. Their input was always on point. Mark Neerincx, thank you for sharing your extensive experience with me and keeping me focused; always with a smile. Revelation is a word I associate with Pim Haselager, never too shy to share his insights. Where would I be without your direct yet always constructive feedback?

I would never have started my internship at TNO without Khiet Truong. She not only introduced me at TNO, she supervised the starting phase of my internship. I am honored that you are the reviewer of my thesis. You were there at the start and you are here now at the end.

I am very thankful for the great time I had during my internship at TNO Soesterberg. All my different skills, knowledge and experiences all came together for the first time. I greatly appreciate how serious my contribution was taken by the staff members. I would especially like to thank Rosemarijn Looije for settling me in, making me feel at home right from the start, and providing me with all the information and resources that I needed. Furthermore, I would like to thank the rest of the PAL crew, Olivier Blanson Henkemans, Sylvia van der Pal, Bert Bierman, Rifca Peters and Frank Kaptein, for the cheerful cooperation.

This internship would not have been such a rich experience without the user study at hospital de Gelderse Vallei. Letting children with diabetes actually use myPAL for three weeks was not only scientifically relevant it was the first time for me that someone actually used and benefited from something I created. I would like to thank Gert Jan van der Burg and Marian van IJzendoorn for enabling me to do that.

Lastly, I would like to thank Elvira Visser. You not only cut me some slack in our household choirs when I was working on this thesis in the weekends, you were there for me whenever I needed support. When I say support I mean that in the broadest sense possible. You instructed me to keep writing and you said it was alright to take a break. You took the lovely pictures for myPAL and you even helped create dialogue items for the avatar. And for that I am eternally grateful.

9. References

[1] International Diabetes Federation, "IDF Diabetes Atlas, 7th ed." 2015.

[2] R. Looije, "Personal Assistant for a healthy Lifestyle." 2014, www.pal4u.eu.

[3] E. Arsand, D. H. Frøisland, S. O. Skrøvseth, T. Chomutare, N. Tatara, G. Hartvigsen, and J. T. Tufano, "Mobile health applications to assist patients with diabetes: lessons learned and design implications.," *J. Diabetes Sci. Technol.*, vol. 6, no. 5, pp. 1197–206, 2012.

[4] M. W. Tsang, M. Mok, G. Kam, M. Jung, A. Tang, U. Chan, C. M. Chu, I. Li, and J. Chan, "Improvement in diabetes control with a monitoring system based on a hand-held, touch-screen electronic diary," *J Telemed Telecare*, vol. 7, no. 1, pp. 47–50, 2001.

[5] E. van der Drift, R. Beun, R. Looije, and O. Henkemans, "A remote social robot to motivate and support diabetic children in keeping a diary," in *Proceedings of Human-Robot Interaction 2014*, 2014, no. Figure 1, pp. 463–470.

[6] J. Tran, R. Tran, and J. R. White, "Smartphone-based glucose monitors and applications in the management of diabetes: An overview of 10 salient 'apps' and a novel smartphone-connected blood glucose monitor," *Clin. Diabetes*, vol. 30, no. 4, pp. 173–178, 2012.

[7] T. Overview and M. E. Therapy, "Invited commentary motivating children and adolescents to improve adherence," *J. Pediatr.*, vol. 3476, no. 03, pp. 430–433, 2003.

[8] K. I. Rother, "Diabetes treatment—bridging the divide," *N. Engl. J. Med.*, vol. 356, no. 15, p. 1499, 2007.

[9] World Health Organization, "Diabetes Fact sheet," 2011.

[10] R. R. Rubin and M. Peyrot, "Psychological issues and treatments for people with diabetes," *J. Clin. Psychol.*, vol. 57, no. 4, pp. 457–478, 2001.

[11] B. Garrofé, A. Björnberg, and A. Phang, "Euro diabetes index." Health Consumer Powerhouse, 2014.

[12] NICE, "Depression in Children and Young People," in *The Medical journal of Australia*, vol. 5, no. 1, Springer, 2005, p. 63.

[13] K. D. Barnard, C. E. Lloyd, and T. C. Skinner, "Systematic literature review: quality of life associated with insulin pump use in Type 1 diabetes," *Diabet Med*, vol. 24, no. 6, pp. 607–617, 2007.

[14] H. KK, S. Huestis, A. Maher, D. Butler, L. Volkening, and L. LMB, "Depressive symptoms in children and adolescents with type 1 diabetes: association with diabetes-specific characteristics.," *Diabetes Care*, vol. 29, no. 6, pp. 1389–1391 3p, 2006.

[15] R. Pfeifer and C. Scheier, *Understanding intelligence*. 1999.

[16] M. Ligthart and K. P. Truong, "Selecting the right robot: Influence of user attitude, robot sociability and embodiment on user preferences," in *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 2015, vol. 2015-Novem, pp. 682–687.

[17] K. Dautenhahn and A. Billard, "Bringing up robots or—the psychology of socially intelligent robots: From theory to implementation," in *Proceedings of the third annual conference on Autonomous Agents*, 1999, pp. 366–367.

[18] C. Breazeal, “Toward sociable robots,” *Rob. Auton. Syst.*, vol. 42, no. 3, pp. 167–175, 2003.

[19] T. Fong, I. Nourbakhsh, and K. Dautenhahn, “A survey of socially interactive robots,” *Rob. Auton. Syst.*, vol. 42, no. 3, pp. 143–166, 2003.

[20] B. R. Duffy, “Anthropomorphism and the social robot,” *Rob. Auton. Syst.*, vol. 42, no. 3, pp. 177–190, 2003.

[21] C. Bartneck and J. Forlizzi, “A design-centred framework for social human-robot interaction,” in *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication*, 2004, pp. 591–594.

[22] P. H. Kahn, N. G. Freier, B. Friedman, R. L. Severson, and E. N. Feldman, “Social and moral relationships with robotic others?,” in *Robot and Human Interactive Communication, 2004. ROMAN 2004. 13th IEEE International Workshop on*, 2004, pp. 545–550.

[23] D. Feil-Seifer and M. J. Mataric, “Defining socially assistive robotics,” in *Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on*, 2005, pp. 465–468.

[24] J. E. Young, *Exploring social interaction between robots and people*. University of Calgary, 2010.

[25] S. M. Rabbitt, A. E. Kazdin, and B. Scassellati, “Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use,” *Clin. Psychol. Rev.*, vol. 35, pp. 35–46, 2015.

[26] R. M. Agrigoroaie and A. Tapus, “Developing a Healthcare Robot with Personalized Behaviors and Social Skills for the Elderly,” in *The Eleventh ACM/IEEE International Conference on Human Robot Interation*, 2016, pp. 589–590.

[27] J. Fasola and M. J. Mataric, “Using socially assistive human--robot interaction to motivate physical exercise for older adults,” *Proc. IEEE*, vol. 100, no. 8, pp. 2512–2526, 2012.

[28] R. Simut, C. A. Costescu, J. Vanderfaeillie, G. van de Perre, B. Vanderborght, D. Lefeber, and others, “Can You Cure me? Children With Autism Spectrum Disorders Playing a Doctor Game With a Social Robot,” *Int. J. Sch. Heal.*, no. Inpress, 2016.

[29] L. Boccanfuso, S. Scarborough, R. K. Abramson, A. V Hall, H. H. Wright, and J. M. O’Kane, “A low-cost socially assistive robot and robot-assisted intervention for children with autism spectrum disorder: field trials and lessons learned,” *Auton. Robots*, pp. 1–19, 2016.

[30] S. M. Rabbitt, A. E. Kazdin, and J. H. Hong, “Acceptability of robot-assisted therapy for disruptive behavior problems in children.,” *Arch. Sci. Psychol.*, vol. 3, no. 1, p. 101, 2015.

[31] T. Belpaeme, P. Baxter, J. De Greeff, J. Kennedy, R. Read, R. Looije, M. Neerincx, I. Baroni, and M. C. Zelati, *Child-robot interaction: Perspectives and challenges*. Springer, 2013.

[32] R. Ros, M. Nalin, R. Wood, P. Baxter, R. Looije, Y. Demiris, T. Belpaeme, A. Giusti, and C. Pozzi, “Child-robot interaction in the wild: advice to the aspiring experimenter,” in *Proceedings of the 13th international conference on multimodal interfaces*, 2011, pp. 335–342.

[33] M. O’Reilly, N. Dogra, and P. D. Ronzoni, *Research with children: Theory and practice*. 2013.

[34] N. Sharkey and A. Sharkey, “The crying shame of robot nannies: an ethical appraisal,” *Interact. Stud.*, vol. 11, no. 2, pp. 161–190, 2010.

[35] L. D. Riek and D. Howard, “A code of ethics for the human-robot interaction profession,” in *Proceedings of We Robot*, 2014.

[36] M. A. Neerincx, “situated Cognitive Engineering.” 2012.

[37] H. a Simon, *The sciences of the artificial, (third edition)*, vol. 33, no. 5. MIT press, 1997.

[38] W. Karwowski, *International Encyclopedia of Ergonomics and Human Factors, Second Edition*, vol. 3. Crc Press, 2006.

[39] K. M. Wilson, W. S. Helton, and M. W. Wiggins, “Cognitive engineering,” *Encycl. Softw. Eng.*, 2012.

[40] J. S. Brown, A. Collins, and P. Duguid, “9 Situated cognition and the culture of learning,” *Educ. Res.*, vol. 18, no. 1, pp. 32–42, 1989.

[41] M. Heidegger, “Die frage nach der technik,” *Vor. und Aufsätze*, vol. 6, no. 3, pp. 9–41, 1990.

[42] M. A. Neerincx and J. Lindenberg, “Situated cognitive engineering for complex task environments.,” *Nat. Decis. Mak. Macrocognition*, pp. 373–390, 2008.

[43] R. Looije and M. A. Neerincx, “Situated Cognitive Engineering to Support Systematic Study of ‘ Social Features ’ in HRI,” 2012.

[44] TUD and TNO, “Applying the Situated Cognitive Engineering Method - A Comprehensive Guide,” no. 1, 2013.

[45] E. J. G. Van Der Drift, “Focus group summary in Master Thesis,” 2014.

[46] R. Ryan and E. Deci, “Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions.,” *Contemp. Educ. Psychol.*, vol. 25, no. 1, pp. 54–67, 2000.

[47] E. L. Deci and R. M. Ryan, “The ‘ What ’ and ‘ Why ’ of Goal Pursuits: Human Needs and the Self-Determination of Behavior,” *Psychol. Inq.*, vol. 11, no. 4, pp. 227–268, 2000.

[48] E. L. Deci and R. M. Ryan, “Self-Determination Theory Home page,” *Selfdeterminationtheory.Org*, 2016. [Online]. Available: http://selfdeterminationtheory.org/.

[49] R. Ryan and E. Deci, “Self-determination theory and the facilitation of intrinsic motivation,” *Am. Psychol.*, vol. 55, no. 1, pp. 68–78, 2000.

[50] S. J. Sebire, R. Jago, K. R. Fox, M. J. Edwards, and J. L. Thompson, “Testing a self-determination theory model of children’s physical activity motivation,” *Int. J. Behav. Nutr. Phys. Act.*, vol. 10, no. 1, p. 111, 2013.

[51] M. Joussemet, R. Koestner, N. Lekes, and N. Houlfort, “Introducing Uninteresting Tasks to Children: A Comparison of the Effects of Rewards and Autonomy Support,” *J. Pers.*, vol. 72, no. 1, pp. 139–166, 2004.

[52] T. C. Society, “The Good Childhood Report 2013,” pp. 1–54, 2013.

[53] I. Katz and A. Assor, “When choice motivates and when it does not,” *Educ. Psychol. Rev.*, vol. 19, no. 4, pp. 429–442, 2007.

[54] C. a. Rohrbeck, M. D. Ginsburg-Block, J. W. Fantuzzo, and T. R. Miller, “Peer-assisted learning interventions with elementary school students: A meta-analytic review.,” *J. Educ. Psychol.*, vol. 95, no. 2, pp. 240–257, 2003.

[55] C. R. Stefanou, K. C. Perencevich, M. DiCintio, and J. C. Turner, “Supporting Autonomy in the Classroom: Ways Teachers Encourage Student Decision Making and Ownership,” *Educ. Psychol.*, vol. 39, no. 2, pp. 97–110, 2004.

[56] R. Koestner, M. Zuckerman, and J. Koestner, “Praise, involvement, and intrinsic motivation.,” *J. Pers. Soc. Psychol.*, vol. 53, no. 2, p. 383, 1987.

[57] A. J. Elliot and H. a McGregor, “A 2x2 achievement goal framework,” *J. Pers. Soc. Psychol.*, vol. 80, no. 3, pp. 501–519, 2001.

[58] D. W. Mood, J. E. Johnson, and C. U. Shantz, “Social comprehension and affect-matching in young children,” *Merril- Palmer Q.*, vol. 24, no. 1, pp. 63–66, 1978.

[59] O. Blanson-Henkemans, Looije, L. Alpay, Neerincx, Hoondert, and Groot, ““I just have diabetes”: children’s need for diabetes self-management support and how a social robot can accommodate their needs,” *Patient Intell.*, vol. 4, p. 51, 2012.

[60] O. A. Blanson Henkemans, B. P. B. Bierman, J. Janssen, M. A. Neerincx, R. Looije, H. van der Bosch, and J. A. M. van der Giessen, “Using a robot to personalise health education for children with diabetes type 1: A pilot study.,” *Patient Educ. Couns.*, vol. 92, no. 2, pp. 174–181 8p, 2013.

[61] M. Langheinrich, “Privacy by design—principles of privacy-aware ubiquitous systems,” in *Ubicomp 2001: Ubiquitous Computing*, 2001, pp. 273–291.

[62] M. Toure-Tillery and A. Fishbach, “How to Measure Motivation: A Guide for the Experimental Social Psychologist,” *Soc. Personal. Psychol. Compass*, vol. 8, no. 7, pp. 328–341, 2014.

[63] S. M. Fulmer and J. C. Frijters, “A review of self-report and alternative approaches in the measurement of student motivation,” *Educ. Psychol. Rev.*, vol. 21, no. 3, pp. 219–246, 2009.

[64] H. Aarts, A. Dijksterhuis, and P. De Vries, “On the psychology of drinking: Being thirsty and perceptually ready,” *Br. J. Psychol.*, vol. 92, no. Pt 4, pp. 631–642, 2001.

[65] J. F. Kroll and M. C. Potter, “Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions,” *J. Verbal Learning Verbal Behav.*, vol. 23, no. 1, pp. 39–66, 1984.

[66] M. J. Ferguson and J. A. Bargh, “Liking is for doing: the effects of goal pursuit on automatic evaluation.,” *J. Pers. Soc. Psychol.*, vol. 87, no. 5, pp. 557–572, 2004.

[67] A. G. Greenwald, D. E. McGhee, and J. L. . Schwartz, “Measuring individual differences in implicit cognition: the implicit association test.,” *J. Pers. Soc. Psychol.*, vol. 74, no. 6, pp. 1464–1480, 1998.

[68] R. M. Ryan, “Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory.,” *J. Pers. Soc. Psychol.*, vol. 43, no. 3, pp. 450–461, 1982.

[69] R. M. Ryan, R. Koestner, and E. L. Deci, “Ego-involved persistence: When free-choice behavior is not intrinsically motivated,” *Motiv. Emot.*, vol. 15, no. 3, pp. 185–205, 1991.

[70] E. McAuley, T. Duncan, and V. V Tammen, “Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: a confirmatory factor analysis.,” *Res. Q. Exerc. Sport*, vol. 60, no. 1, pp. 48–58, 1989.

[71] D. Read, G. Loewenstein, and S. Kalyanaraman, “Mixing virtue and vice: Combining the immediacy effect and the diversification heuristic,” *J. Behav. Decis. Mak.*, vol. 12, no. 4, pp. 257–273, 1999.

[72] UXPin, *Web UI Design for the Human Eye*. 2015, (e-book: https://studio.uxpin.com/ebooks/visual-web-ui-design-colors-space-contrast/).

[73] G. Norman, “Likert scales, levels of measurement and the ‘laws’ of statistics,” *Adv. Heal. Sci. Educ.*, vol. 15, no. 5, pp. 625–632, 2010.

[74] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine Learning in {P}ython,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[75] R. L. Thorndike, “Who belongs in the family?,” *Psychometrika*, vol. 18, no. 4, pp. 267–276, 1954.

[76] D. J. Ketchen and C. L. Shook, “The application of cluster analysis in strategic management research: an analysis and critique,” *Strateg. Manag. J.*, vol. 17, no. 6, pp. 441–458, 1996.

[77] E. M. Segura, M. Kriegel, R. Aylett, A. Deshmukh, and H. Cramer, “How do you like me in this: User embodiment preferences for companion agents,” in *Intelligent Virtual Agents*, 2012, pp. 112–125.

[78] D. George and M. Mallery, “Using SPSS for Windows step by step: a simple guide and reference,” 2003.

[79] G. C. Williams, V. M. Grow, Z. R. Freedman, R. M. Ryan, and E. L. Deci, “Motivational predictors of weight loss and weight-loss maintenance.,” *J. Pers. Soc. Psychol.*, vol. 70, no. 1, p. 115, 1996.

[80] T. Kasser, J. Davey, and R. M. Ryan, “Motivation and employee-supervisor discrepancies in a psychiatric vocational rehabilitation setting.,” *Rehabil. Psychol.*, vol. 37, no. 3, p. 175, 1992.

[81] G. C. Williams, Z. R. Freedman, and E. L. Deci, “Supporting autonomy to motivate patients with diabetes for glucose control,” *Diabetes Care*, vol. 21, no. 10, pp. 1644–1651, 1998.

[82] K. M. Sheldon, R. Ryan, and H. T. Reis, “What makes for a good day? Competence and autonomy in the day and in the person,” *Personal. Soc. Psychol. Bull.*, vol. 22, pp. 1270–1279, 1996.

[83] C. S. Levesque, G. C. Williams, D. Elliot, M. A. Pickering, B. Bodenhamer, and P. J. Finley, “Validating the theoretical structure of the Treatment Self-Regulation Questionnaire (TSRQ) across three different health behaviors,” *Health Educ. Res.*, vol. 22, no. 5, pp. 691–702, 2007.

A. Appendix

1. Interactional dataset

1.1. User behavior

The data set for user behavior contains: a log ID, a timestamp, a type ID and a user name. The type ID referred to an item in the database containing the following user behavior types. The ones marked with an asterisk (*) were used to construct the interaction profiles per user.

Overhead

Login*
Logoff*
Access calendar
Update calendar down: navigate back in time
Update calendar up: navigate forward in time
Update calendar directly: manually select date

Activities

Together or alone*: push add activity button
Access add activity page: selected 'alone'
Together*: selected 'together'
Added activity*

View activity*
Access update activity page
Update activity*
Delete activity*

Access add activity type page
Added activity type*
Delete activity type*

Measurements

Access select measurement page*
Access add glucose page
Added glucose*
Access add insulin page
Added insulin*

View measurement
Access update glucose page
Updated glucose*
Access update insulin page
Updated insulin*

Delete glucose*
Delete insulin*

Added with Gluconline

Pictures

Select picture from gallery page
Link picture to activity*
Access add picture directly page
Added picture directly*

Access gallery*
Access add picture page
Added picture*

Delete picture from activity*
Unlink picture from activity*
Delete picture from gallery*

Goals

Access goals*

Access goal add daily page
Added goal daily*

Access goal add total page
Added goal total*

Delete goal*

1.2. Avatar behavior

The avatar behavior dataset consists of a log ID, a timestamp, a type ID and a user name. The type ID referred to an item in the database containing the following avatar behavior types:

- Greeting
- Ask (to add an activity) Together or Alone, Add activity together, Add activity alone
- React Happy School, React Neutral School, React Sad School,
- React Happy Sport, React Neutral Sport, React Sad Sport,
- React Happy Meal, React Neutral Meal, React Sad Meal,
- React Happy Other, React Neutral Other, React Sad Other
- Show Picture
- Encourage Adding daily/total Goal, React daily/total Goal Added
- Encourage Active daily/total Goal, Complimented Met daily/total Goal

1.3. User added content

MyPAL stores the following user added content:

Activities containing:

- Activity ID
- User Name
- Added (Timestamp)
- Date
- Start time
- End time
- Activity Type
- Activity Name
- Description
- Emotion
- Picture ID
- Carbohydrate value (= -1.0 if activity type is not a meal)

Activity types (users could add custom types besides the default school, meal, sport and other) containing:

- ID
- User Name
- Type Name
- Color
- Icon

Pictures (the actual pictures are stored on disk and a pointer is stored in the database. This refers to the pointer) containing:

- ID
- User Name
- Added (Timestamp)
- Date
- Diary Activity ID (if linked)
- File Path
- Thumbnail Path

Glucose and **Insulin Measurements** (stored separately) containing:

- Measurement ID
- User Name
- Added (Timestamp)
- Date
- Start time
- End time
- Day Part
- Value
- Comment

Goal containing:

- ID
- User Name
- Added (Timestamp)
- Goal Type (Daily or Total goal)
- Target
- Target Value (For daily between 1 and 10; for Total between 2 and 7)
- Start Date
- Deadline
- isMet (Boolean)
- is met at (Timestamp)

The target corresponds with the following targets:

ADDxACTIVITIES, ADDxMEASUREMENTS, ADDxPICTURES, ADDxYESTERDAY, CONACTIVITIES, CONMEASUREMENTS, CONPICTURES, CONLOGINS;

2. Questions dataset

At three moments during the user study questionnaires measured various constructs: during the introduction session, halfway the three week use of myPAL and during the evaluation session. With a Cronbach's Alpha analysis, the reliability for each SDT-construct was investigated. If all the combined items have a score of 0.7 or higher they reliably represent the same construct [78]. If the combined items did not reach that threshold the questions were inspected more closely. Is there one question that causes the reliability to drop or are there more? Another Cronbach's Alpha was calculated leaving out the interfering items. If a score of 0.7 or higher was obtained that subset of items was used to represent the construct instead. The SDT constructs are represented by an average score of all the included items. The non-SDT constructs are evaluated separately. Therefore, a reliability analysis is not necessary.

Another step is taken to validate the quality of the SDT constructs. Three antecedents of intrinsic motivation are measured together with an overall estimation of the motivation of using myPAL. It is to be expected that the antecedents correlate with the general measure of motivation. The check this a correlation analysis is performed for these constructs at each point of measurement.

The items of the constructs measuring the feelings of autonomy, competence and relatedness and the general intrinsic motivation are adapted from validated SDT questionnaires used in [70], [79]–[83]. The other questions were created especially for this research to either check the expectations the participants had beforehand and whether the manipulations were noticed properly and to evaluate the avatar and myPAL afterwards.

Furthermore, during the introduction and evaluation sessions the participants were interviewed. The questions are also listed below.

2.1. Introduction session

The following items are translated from the original Dutch questionnaire. The questions marked with an * where used in the interview as well.

Feeling of autonomy

- I'm going to keep this diary because: I have to – I want it. *
- I participate in this experiments because: my parents want me to – I decided to do so myself.

Cronbach's alpha of .26. Because the other autonomy questions were directed at the diary and not the experiment the second question was dropped.

Feeling of competence

- Keeping a diary is something I: find difficult – am well able to do. *
- Keeping a promise is something I find: difficult – easy to do.

Cronbach's alpha of -.26. The competence required for keeping a diary and keeping promises is presumably of a different kind. It any case, the questions do not seem to reliably measure the same construct. Because the other competence questions are mostly directed at the diary or other myPAL functionalities the question about promise was dropped.

Feeling of relatedness

- Becoming friends with a robot is something: I cannot do – I can do. *

- Talking with the robot is something I want to do: as little as possible – as much as possible.

Cronbach's alpha of -.88. These questions do not align at all. The question about friendship is dropped because the other relatedness questions are about interacting with the robot and avatar not about concepts of friendship.

General intrinsic motivation

- I thing that I will find the diary: boring – fun to use.
- In the diary I will put not much – a lot of effort.
- To keep a diary is not important – very important.
- I think that this diary will not help – help me to deal with my diabetes.

Cronbach's alpha of -.40. Removing the second question has a significant increase in reliability (alpha of .64). The second question was dropped. None of the antecedents significantly correlated with the general measure of intrinsic motivation:

	Autonomy	Competence	Relatedness
Pearson's r (p-value)	-.010 (.973)	.469 (.106)	-.071 (.817)

Experience with computers

- I use the computer / tablet / mobile phone seldom - often
- I play games on the computer / tablet / mobile phone seldom – often

Expectations of using myPAL (Interview)

- What do you think you should write in a diabetes diary?
- How often do you think you need to update your diary?
- Give an estimation of the time it takes to update a diary?

Experience with robots and avatars (Interview)

- Name all the robots from movies, books, games, real life, etc. that you know.
- Are you familiar with this robot (see Figure 65)?

Figure 65: image of Robin the Robot show to children during the interview

- If yes, how do you know this robot?
- Can you explain what an 'avatar' is?
- Did you ever have a conversation with an avatar via the computer or otherwise?
- If yes, what was the conversation about?

Expectations of interaction with the avatar (Interview)

- What can you do with an avatar?
- What can you discuss with an avatar?

2.2. Halfway

Feeling of autonomy

- I keep this diary because I have to – I want to
- When I log in I do that because my parents remind me – I think of it myself.

Cronbach's alpha of .74.

Feeling of competence

- Adding activities and measurements is something I find hard – easy
- Meeting the goals is something I find hard – easy to do

Cronbach's alpha of .79.

Feeling of relatedness

- If Robin wasn't present in the diary I would add less – more content
- Everything Robin says I don't like – like
- Everything Robin says is useless – useful

Cronbach's alpha of .89. A significant increase can be obtained by removing the first item (alpha of .93). The first question is removed to keep the questionnaire consistent with the final questionnaire were this item was removed to gain enough reliability for the construct.

General intrinsic motivation

- I consider the diary to be boring – fun
- For keeping the diary, I invest not any effort – much effort
- Keeping the diary is something I find unimportant – important
- The diary does not – does help me to deal with my diabetes

Cronbach's alpha of .88. A small increase (.89) can be obtained by dropping the second question. To conserve the consistency between the pre and post questionnaires that question was dropped. None of the antecedents significantly correlated with the general measure of intrinsic motivation. Autonomy however significantly correlated with relatedness ($r = .671$; $p = .024$).

	Autonomy	Competence	Relatedness
Pearson's *r* (*p*-value)	.573 (.065)	.457 (.157)	-.421 (.197)

2.3. Evaluation session

The questions marked with an * where used in the interview as well.

Feeling of autonomy

- I kept this diary because I have to – I want to*

- When I logged in I do that because my parents remind me – I think of it myself.

Cronbach's alpha of .76.

Feeling of competence

- Adding activities is something I found hard – easy to do*
- Adding measurements is something I found hard – easy to do*
- Adding pictures is something I found hard – easy to do
- Adding goals is something I found hard – easy to do
- Keeping goals is something I found hard – easy to do

Cronbach's alpha of .60. Removing one the questions does not seem to improve the reliability much. The difference with the half-way questionnaire is that this time each functionality is reviewed separately. Some participants reported to find some parts of myPAL harder than others. It makes sense that the reliability as a whole is affected by a larger variance in the answers. Regardless all the questions are kept in to calculate one average because it represents the overall feeling of competence best. A Cronbach's alpha of .60 is considered questionable but not poor. This is a factor to keep in mind while performing the analyses.

Feeling of relatedness

- If Robin wasn't present in the diary I would add less – more content
- Everything Robin said I didn't like – like
- Everything Robin said was useless – useful

Cronbach's alpha of .58 which is considered poor. A significant improvement would be to remove the first questions (alpha of .78). Conceptually this makes sense because it asks about the influence of the avatar instead of evaluating the behavior of the avatar. The same improvement by removing the first question is also observed with the feeling of relatedness construct of the half-way questionnaire. Only the last two questions were used for the analyses.

General intrinsic motivation

- I considered the diary to be boring – fun
- For keeping the diary, I invested not any effort – much effort
- Keeping the diary was something I found unimportant – important
- The diary did not – did help me to deal with my diabetes
- If I could continue using the diary I wouldn't – would*

Cronbach's alpha of .72. The second question is dropped to remain consistent with the pre and half-way questionnaires. This results in an alpha of .84. A strong correlation has been found between the relatedness and general measure of intrinsic motivation. No further significant correlations were found.

	Autonomy	Competence	Relatedness
Pearson's *r* (*p*-value)	.257 (.397)	.165 (.591)	**.705** (.007)

Avatar behavior check

- Robin did not – did offer to add an activity together

- Robin never – always responded to me when I added an activity
- Robin's response never – always matched with the activity
- Robin's response never – always matched with my indicated mood
- Robin never – always responded to me when I added a measurement
- Robin never – always shared a picture of himself[4]?
- Robin never – always responded to me when I visited the goal page
- Robin never – always responded to me when I added a goal
- Robin never – always responded to me when I completed a goal

Avatar behavior: self-disclosure

- Robin shared little – much about himself (perceived)
- When Robin talked about himself I didn't – did like it* (preferred)
- The pictures Robin shared I didn't – did like (preferred)

Avatar behavior: general responsiveness

- Robin didn't – did understood what I meant (perceived)
- I didn't – did like how robin responded to what I filled in* (preferred)
- I didn't – did like how robin responded to my added activities (preferred)
- I didn't – did like how robin responded to my added goals (preferred)
- I didn't – did like how robin responded to my met goals (preferred)

Avatar behavior: mood matching

- Robin didn't – did understand my feelings (perceived)
- I didn't – did like how Robin responded to my feelings* (preferred)

Avatar behavior: explicit encouragement

- Robin didn't – did encouraged me to add a goal (perceived)
- I didn't – did like how Robin encouraged me* (preferred)

Avatar behavior: praise

- Robin didn't – did praised me (perceived)
- I didn't – did like to be praised by Robin* (preferred)

Self-reported importance of avatar

- Robin wasn't – was important for me to like the diary*
- Robin wasn't – was important for me to log in
- Robin wasn't – was important for me to add activities
- Robin wasn't – was important for me to add measurements
- Robin wasn't – was important for me to add goals
- Robin wasn't – was important for me to meet goals

Evaluation diary use (Interview)

- Can you describe the process of using myPAL at home?
- Did you needed to be reminder often?
- How often did you log in?

---

[4] In Dutch no gender was assigned to Robin in the questions instead 'zichzelf' was used which is a gender neutral self-referencing term.

- Did you added content every time you logged in?
- How long did each session last?

Reliability analysis

The following constructs determine the attitudes towards the avatar:

- Perceived sociability (Cronbach's alpha = .80)
- Behavior preferences (Cronbach's alpha = .83 without question about pictures)
- Self-reported importance (Cronbach's alpha = .81)

The question "the pictures Robin shared I didn't – did like" was removed because it hugely affected the reliability of behavior preference construct (Cronbach's alpha with question is .56). The post interview session revealed that some children did not experience the avatar sharing a picture and rated that particular questions as 'I didn't like' because of it. That is the cause of the found incongruity with the rest of the scores. The pictures question is therefore left out.

The constructs were calculated by taking the mean of all the selected questions. All the constructs follow a normal distribution (Shapiro-Wilk: statistic(13) > .907 and *p*'s > .169) and no (significant) outliers were detected.

Two strong correlations have been found (see Table 4) between the perceived sociability and respectively behavior preference and self-assigned importance. No significant correlation was found between behavior preference and self-assigned importance.

		Perceived sociability	Behavior preference	Self-assigned importance
Perceived sociability	Pearson Correlation	1	,856**	,703**
	Sig. (2-tailed)		,000	,007
	N	13	13	13
Behavior preference	Pearson Correlation	,856**	1	,474
	Sig. (2-tailed)	,000		,102
	N	13	13	13
Self-assigned importance	Pearson Correlation	,703**	,474	1
	Sig. (2-tailed)	,007	,102	
	N	13	13	13

**. Correlation is significant at the 0.01 level (2-tailed).

Table 4: matrix showing the correlations between perceived sociability, behavior preference and self-assigned importance.

3. Avatar behavior (Dutch)

In the following table all the speech elements are listed. Each element can have multiple versions. Each version has an equal probability to be selected by the behavior manager.

Id	Speech
1	(a still image of the avatar)
2	Hoi #firstName#, fijn dat je er bent. Hoi #firstName#, goed dat je er bent. Hoi #firstName#, leuk om je weer te zien. Hey #firstName#, fijn dat je er bent. Hey #firstName#, goed dat je er bent. Hey #firstName#, leuk om je weer te zien. Hallo #firstName#, fijn dat je er bent. Hallo #firstName#, goed dat je er bent. Hallo #firstName#, leuk om je weer te zien.
3	Wil je samen of alleen een activiteit toevoegen?
4	Ik ben benieuwd wat je gedaan hebt. Leuk dat we samen een activiteit gaan invullen. Goed dat je een activiteit gaat toevoegen. Laten we beginnen!
5	Wat voor soort activiteit heb je gedaan? Kun je aanklikken wat je gedaan hebt? Welk type past het beste bij wat je gedaan hebt?
6	Tsjakka, en wat is de naam van de activiteit? Hoppa, en hoe zou je de activiteit willen noemen? Hupsakee, en hoe heet de activiteit?
7	Wanneer was dat? Wanneer heb je dat gedaan? Oké, wanneer vond dit plaats?
8	Van hoe laat tot hoe laat was het? Hoe laat was dat? Hoe laat heb je dat gedaan?
9	Kun je aangeven wat je gevoel daarbij was? Wat was je gevoel daarbij? Hoe voelde je je toen? Wat was je gevoel? Wat zei je gevoel daarover?
10	Kun je er wat meer over vertellen? Vertel eens! Zou je er wat meer over kunnen vertellen? Wat kun je daarover opschrijven? Vertel eens wat meer. Wat wil je er over kwijt?
11	Wil je ook een foto toevoegen?

	Je mag ook een foto toevoegen als je wilt. Doen? Voeg je ook een foto toe?
12	Ben benieuwd! Je kunt de foto hier uploaden. Laat maar komen. Voeg hem hieronder toe. Fijn dat je dat wil doen. Je kunt de foto hieronder uploaden.
13	Prima. Je kunt de activiteit nu toevoegen. Oke, je kunt de activiteit toevoegen. Volgende keer misschien. Je mag de activiteit nu toevoegen.
14	Wat goed! Je hebt een dagdoel behaald. Een dagdoel behaald. Ik ben trots op je! Goed bezig met je dagdoelen, je hebt er eentje behaald.
15	Woehoe, een dagdoel behaald. Vinkje bij een dagdoel. Weer een dagdoel klaar! Goed zo! Een dagdoel behaald.
16	Ja! Een herhaaldoel gehaald. Wow, goed zo! Een herhaaldoel behaald. Ik ben zo trots! Je hebt een herhaaldoel afgevinkt.
17	Dit maakt mij blij! Een herhaaldoel is klaar. Herhaaldoelen zijn het moeilijkst om te halen. En dat is jou gelukt! Super, een herhaaldoel af!
18	Top. Goed dat je een dagdoel hebt toegevoegd. Succes met je dagdoel. Wat een mooi dagdoel. Fijn een dagdoel. Je kunt het.
19	Leuk, een herhaaldoel. Succes met je herhaaldoel. Een herhaaldoel is een goede uitdaging voor je.
20	Ik zie dat er nog een dagdoel actief is, hou vol om deze te halen. Probeer het actieve dagdoel af te ronden. Er is nog een dagdoel actief. Ik weet je het kunt halen!
21	Goed bezig met je dagdoelen, ga zo door. Fijn dat je zo goed bezig bent met je dagdoelen, probeer er nog eentje te halen. Ik zie dat je nog een dagdoel kan afronden, probeer het eens.
22	Als je morgen weer terugkomt kun je het herhaaldoel nog verder afmaken. Kom je morgen weer terug? Dat is goed voor je herhaaldoelen. Weet je nog, bij herhaaldoelen moet je een aantal dagen achter elkaar terugkomen.
23	Herhaaldoelen halen is een kwestie van volhouden. Ik weet dat je het openstaande herhaaldoel kunt halen. Hou vol en je gaat dit herhaaldoel ook halen.
24	Laten we een doel aanmaken. Er zijn geen actieve doelen, tijd om eentje te maken. Zullen we een doel aanmaken?

	Doelen stellen is een goede manier om iets gedaan te krijgen. Maak je er weer eentje?
25	Was het eten vies? Vond je het moeilijk om te eten? Vond je het eten vies?
26	Oh dat is jammer. Wat vervelend om te horen.
27	Gelukkig maar. Wat was er dan?
28	Hopelijk gaat het morgen beter!
29	Was het lekker?
30	Gelukkig, ik houd van lekker eten.
31	Dat is jammer om te horen!
32	Dat kan.
33	Hee, een maaltijd. Wat is eigenlijk je lievelingseten? Hee, een maaltijd. Wat eet je het liefst? Hee, een maaltijd. Wat vind je het lekkerste toetje?
34	Dat is ook de mijne! Ik vind chocolade ijs lekker, maar ik eet vaak toch een appel. Oh dat klinkt erg lekker!
35	Heb je een wedstrijd gespeeld?
36	Ja? Hopelijk heb je gewonnen!
37	Nee? Dan heb je vast getraind of iets anders.
38	Sport je het liefst alleen of samen met anderen?
39	Ik speel het liefste in een team Leuk, Ik voetbal het liefst met mijn robot vriendjes
40	Van sporten word ik blij, jij ook?
41	Fijn!
42	Oke!
43	Niet leuk om te horen dat het sporten wat minder goed ging deze keer. Oh, wat jammer om te zien dat sport minder leuk was. Vervelend dat het sporten niet zo fijn was.
44	Hopelijk gaat het volgende keer beter. Hopelijk is het wat fijner volgende keer. Ik hoop dat het leuker is de volgende keer.
45	Gaaf om te horen dat #activityName# goed ging. Leuk om te horen dat #activityName# je deze keer beviel. Tof dat #activityName# zo leuk was.
46	Kun je me uitleggen wat #activityName# ook al weer is? Ik ben vergen wat #activityName# ook al weer betekent. Wil je dat uitleggen? Kun je iets meer vertellen over #activityName#? Ik zou graag wat meer willen weten van #activityName#.
47	Bedankt dat je iets over #activityName# wil toevoegen. Hee, je een #activityName# activiteit toegevoegd.
48	Is er iets dat het prettiger zou kunnen maken?

	Is er iets wat het leuker zou kunnen maken? Is er iets dat het fijner zou kunnen maken?
49	Tof! Cool! Leuk! Wat fijn dat je school leuk vond! Het was dus leuk op school? Een leuke school op dag gehad, mooi!
50	Dank voor het invullen. Goed dat je naar school bent geweest. School, check! Ik moet soms ook naar school.
51	Dank voor het invullen! Goed dat je weer naar school was. Je vond het dus niet zo leuk op school vandaag? wat jammer! Morgen gaat het hopelijk beter op school.
52	Wat is jouw lievelingsvak op school?
53	Oh ja, dat vind ik ook een leuk vak!
54	Ik ben goed in rekenen, jij ook?
55	Ja?! Hopelijk gaat je volgende toets goed!
56	Nee? Dat is jammer! Vaak oefenen helpt!
57	Wil je mijn foto zien? Vind je het leuk om mijn foto te zien? Ik heb een foto die ik graag met je wil delen. Wil je die zien? Wil je een foto van mij zien?
58	Jammer! Jammer, misschien een andere keer. Helaas, hopelijk de volgende keer wel. Ander keertje dan.
59	Hier was ik aardappeltjes en een worstje aan het eten. Het eten van gisteravond was erg lekker. Ik had wel een beetje geknoeid.
60	Hier was ik op weg naar de bibliotheek om een boek over hamsters te zoeken. Ik had al mijn boeken al uit dus ik ging naar de bieb om nog eentje te lenen.
61	Gisteren was ik heel moe. Toen heb ik de hele avond om te bank gelegen. Soms doe ik een dutje op de bank.
62	Vanmiddag mocht ik op het hobbelpaard! Dat vind ik zo leuk. Er is een hobbelpaard in het winkelcentrum. Ik heb daar even opgezeten.
63	Kijk, mooie hoed he? Ik had laatst een grappige hoed opgezet.
64	Poeh, ik was moe van het sporten. Sporten is leuk, maar vermoeiend!
65	Kijk ik heb een leuke tekening gemaakt! Ik ben heel trots op de tekening die ik heb gemaakt.
66	Ik had geen zin meer om te lopen dus ik ging er maar bij zitten.

	Ik moest even uitrusten tijdens de wandeling buiten.
67	Ik mag per dag een uurtje achter de computer om een spelletje te spelen. Kijk, zo zit er dan bij als ik ga inloggen in jouw dagboek.
68	Ik lijk erg klein, maar ik vind het leuk om te fietsen. Soms pak ik de fiets om ergens heen te gaan.
69	Hier ben ik huiswerk aan het maken. Heb jij ook huiswerk? Hier ben ik bezig om een opdracht te maken.
70	Druk bezig met mijn huiswerk. Huiswerk moet ook gebeuren.
71	Ik teken graag. Kijk, hier ben ik aan het tekenen.
72	Ik lees graag over robots. Ik ben zo blij met dit boek over robots.
73	Dit is echt een spannend verhaal. Ik kan soms wel uren achter elkaar lezen.
74	Kinderverhalen zijn de beste verhalen. Lees jij ook boeken? Ik vind ze echt geweldig.
75	Ik vind boodschappen doen altijd leuk, want dan mag ik in het wagentje zitten. Ga jij weleens boodschappen doen? Ik mag altijd in het wagentje zitten.
76	Lekker mijn boterham met kaas eten. Dit was mijn ontbijt vanochtend.
77	Zo lig ik erbij als ik ga slapen. Lekker naar bedje toe straks met mijn knuffel.
78	Soms heb ik geen zin om op te staan. Het bed ligt wel erg lekker soms.
79	Ik mocht wat uitzoeken uit de automaat! Ik ben benieuwd wat ik ga krijgen uit de automaat.
80	Wiej, glijden van de glijbaan is leuk. Ik vind de speeltuin altijd leuk. Vooral de glijbaan.
81	Naast het sorteerspel ben ik ook goed in vier op een rij. Heb jij weleens vier op een rij gespeeld. Ik vind dat een erg leuk spel.
82	Mooie autootjes heb ik hè? Op woensdagmiddag speel ik altijd met mijn autootjes.
83	Mijn dino's hebben hoedjes. Ik vind het ook leuk om met mijn dino's te spelen.
84	Ik ben Mike aan het verslaan met een potje kaarten. Naast werken kaart ik ook weleens met Mike.
85	Soms lig ik er zo bij als ik inlog in jouw dagboek. Kijk, ik heb ook een tablet waar ik weleens op mag.
86	Ik vind Tai Chi erg ontspannend. Door Tai Chi oefeningen te doen kom ik heerlijk tot rust.
87	Voor het slapen gaan altijd de tandjes poetsen. Poets jij je tanden wel goed voordat je gaat slapen? Ik wel.
88	Ik vind voetbal kijken ook erg leuk om te doen.

	Naast zelf te voetballen vind ik het ook leuk om het te kijken op het veldje.
89	Goal! Ik doe aan voetballen. Mijn hobby is voetbal. De sport die ik doe is voetbal.

4. Results user study

The results showed in this section are summaries generated per user. For each user an interaction timeline and a motivation timeline is constructed. The interaction timeline covers the behavior of the user and the avatar regarding content management and characteristics of the content. The motivation timeline displays the three measurements of the general motivation and the feeling of autonomy, competence, and relatedness.

4.1. Behavior timeline per user

For each user a timeline is constructed that shows their behavior regarding the management activities, measurements, pictures and goals. The profile furthermore contains an overview of the times the avatar responded to that use. Finally, also information about the added content, such as number of words, the difficulty of the goals and the personal level of the activity descriptions, is included in the profile.

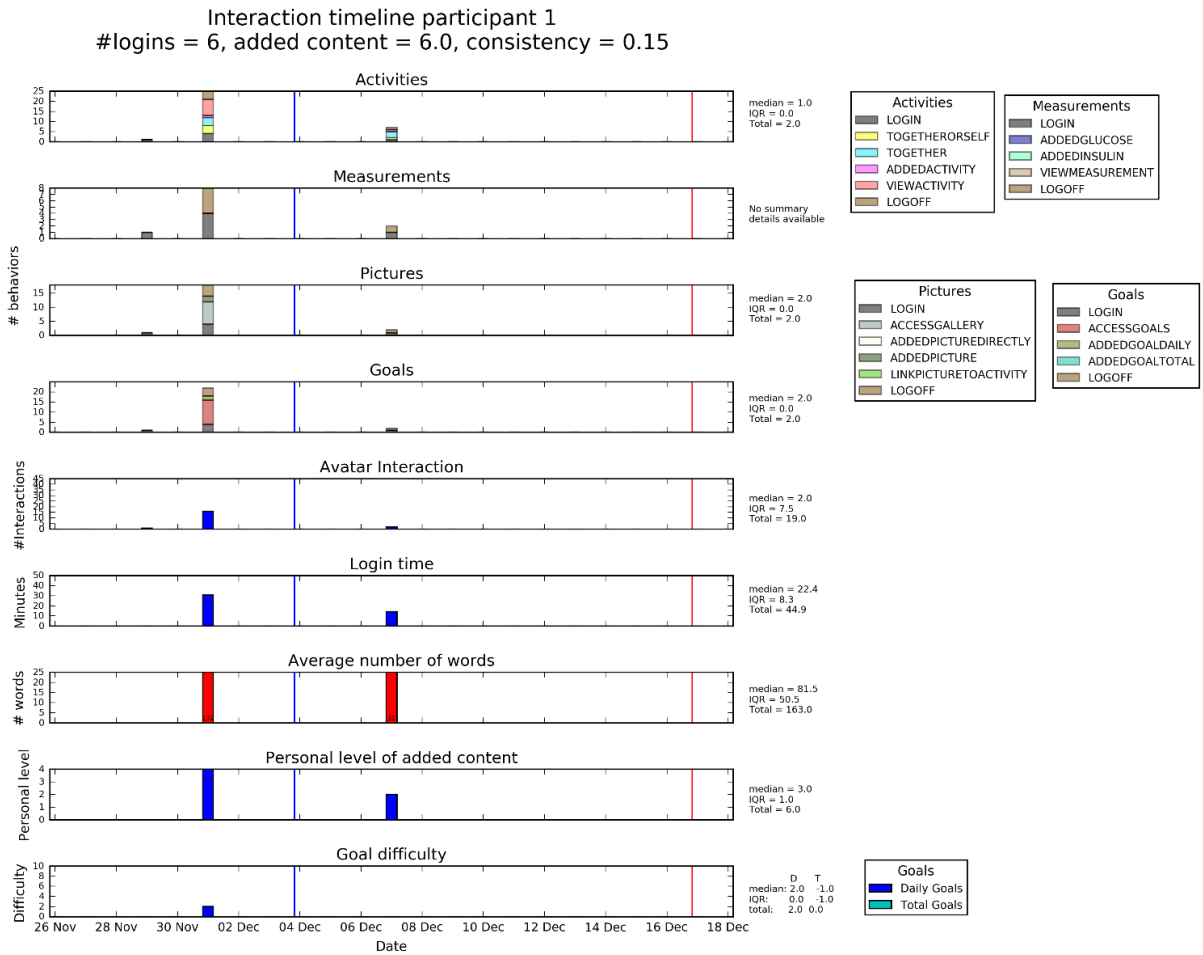


Figure 66: interaction timeline participant 1

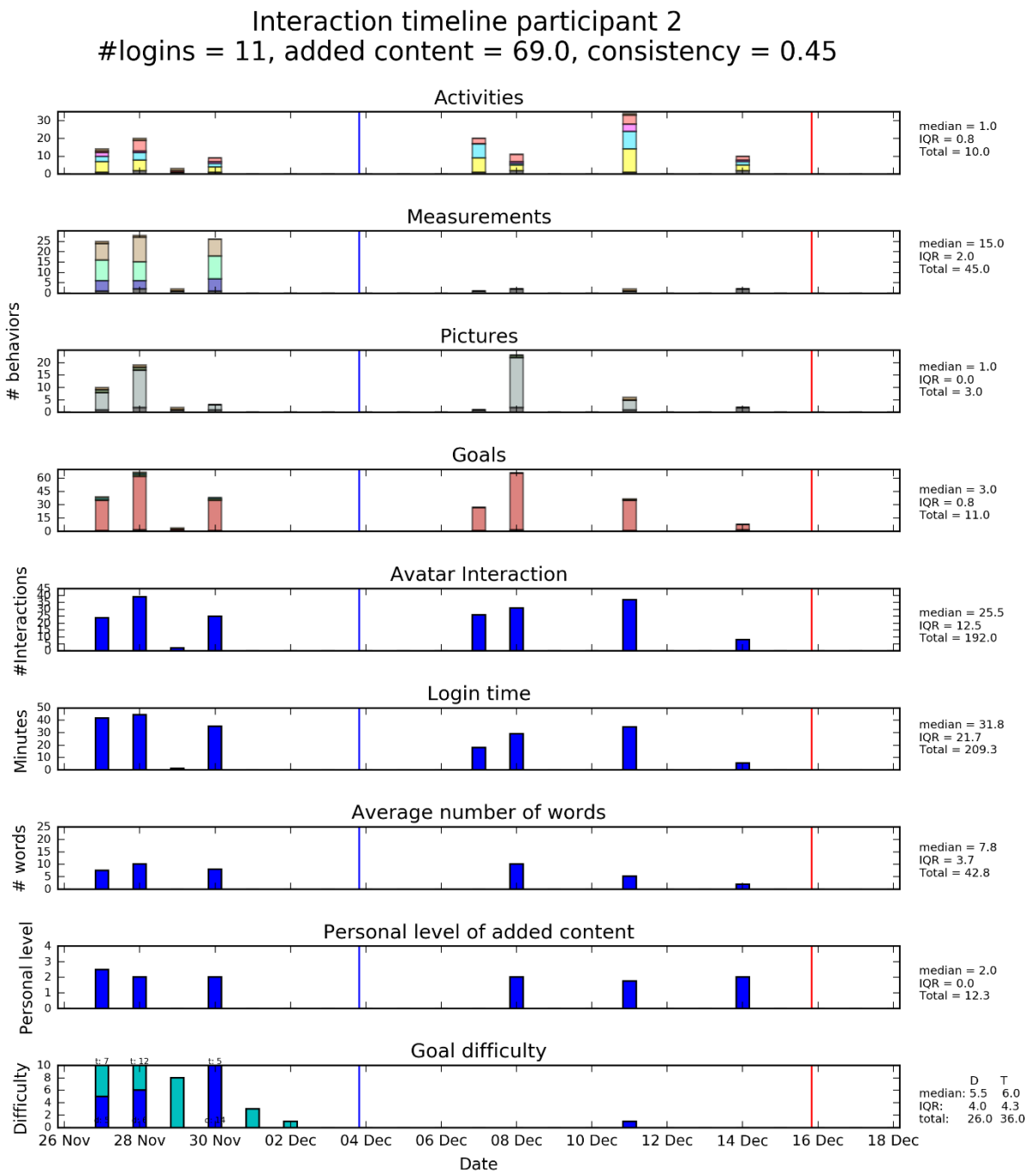


Figure 67: Interaction timeline participant 2

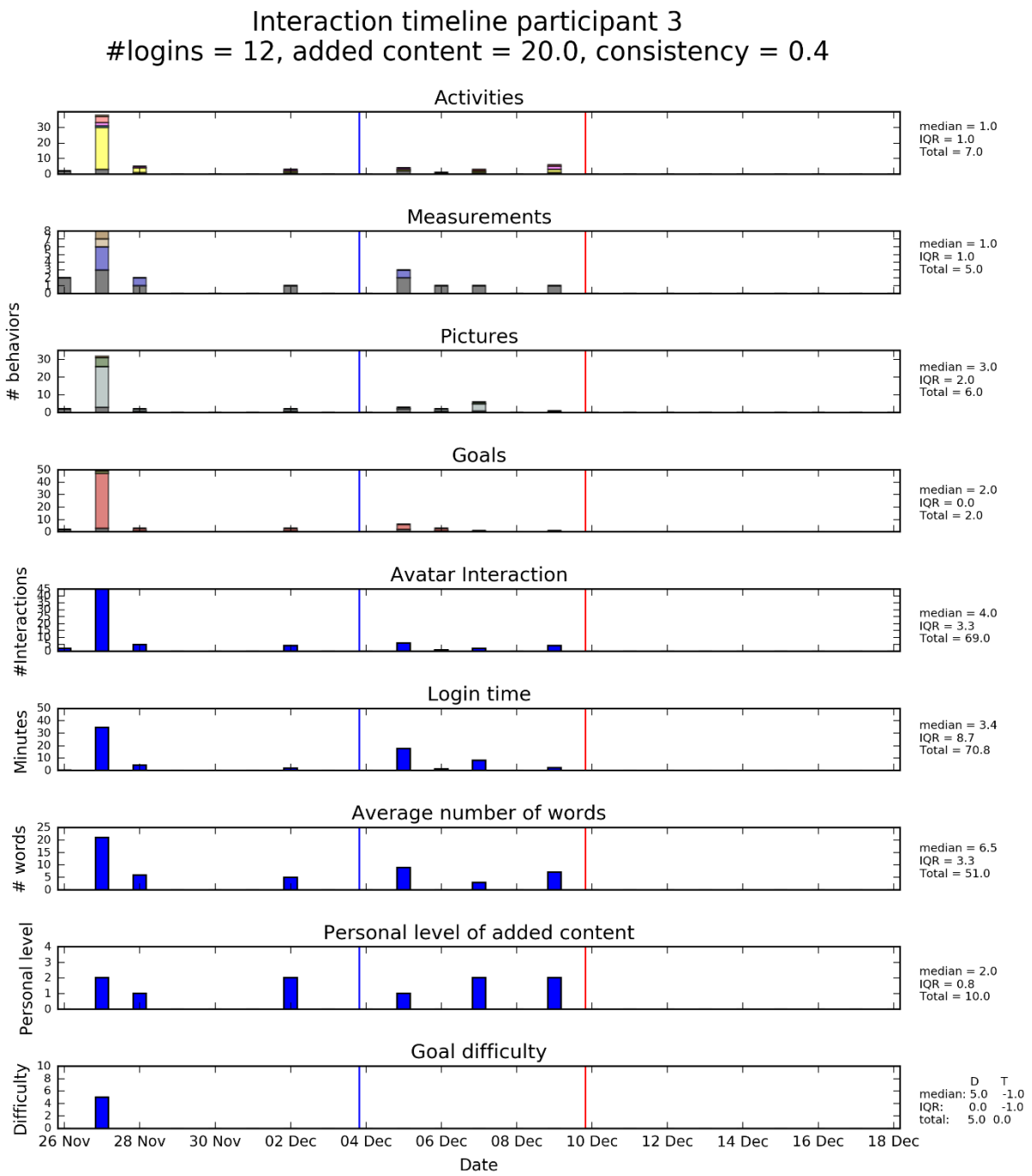


Figure 68: Interaction timeline participant 3

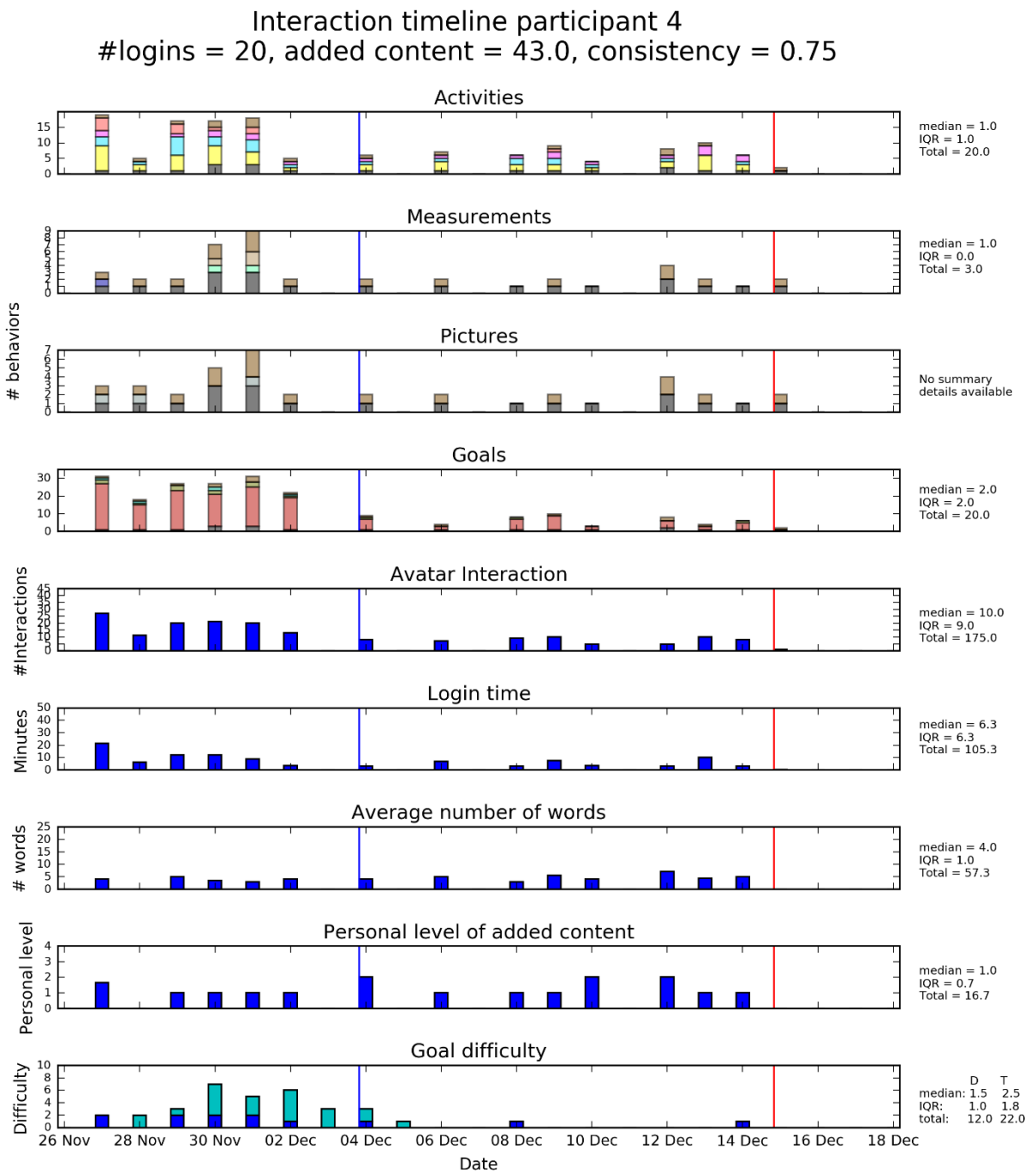


Figure 69: interaction timeline participant 4

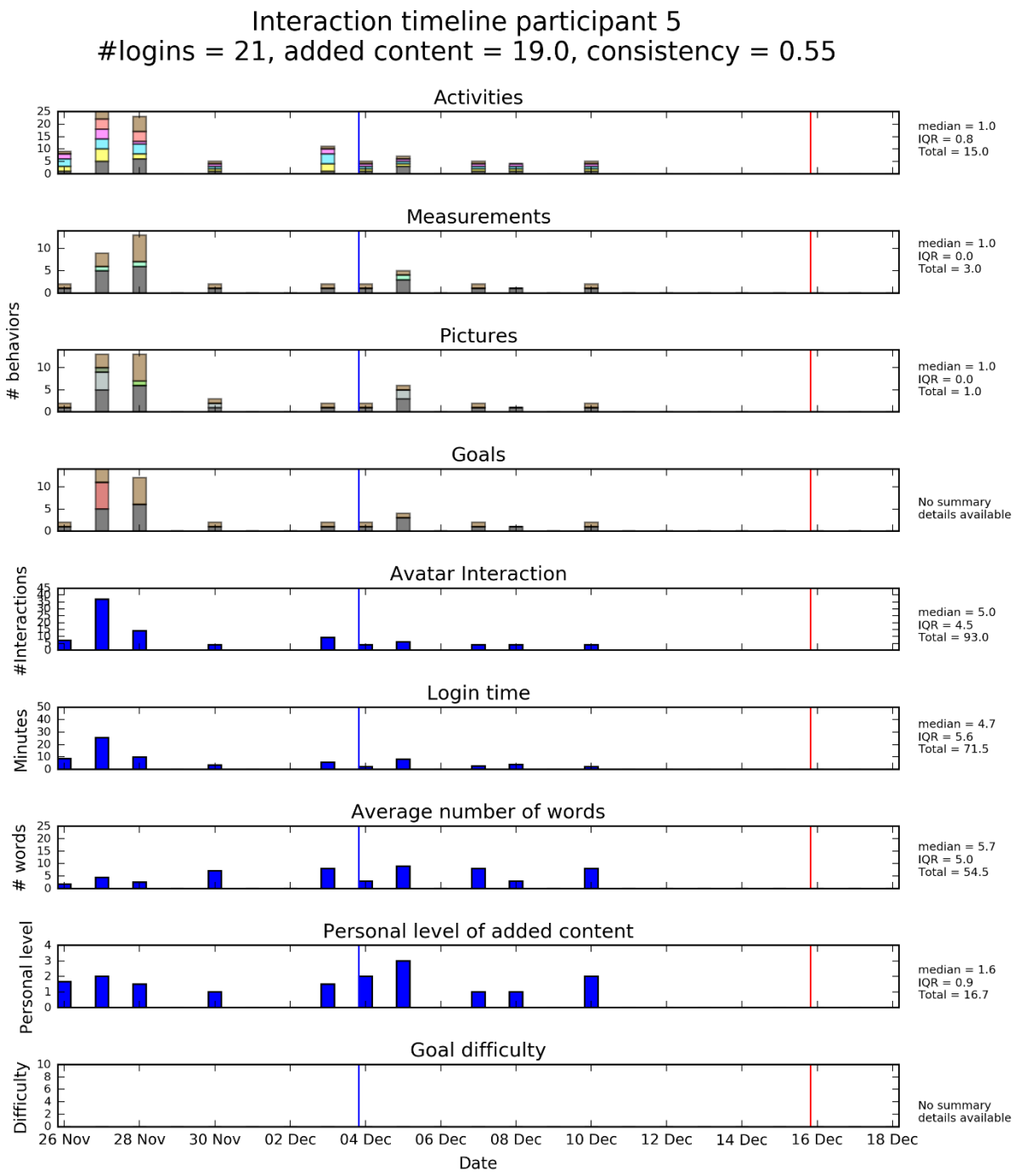


Figure 70: Interaction timeline participant 5

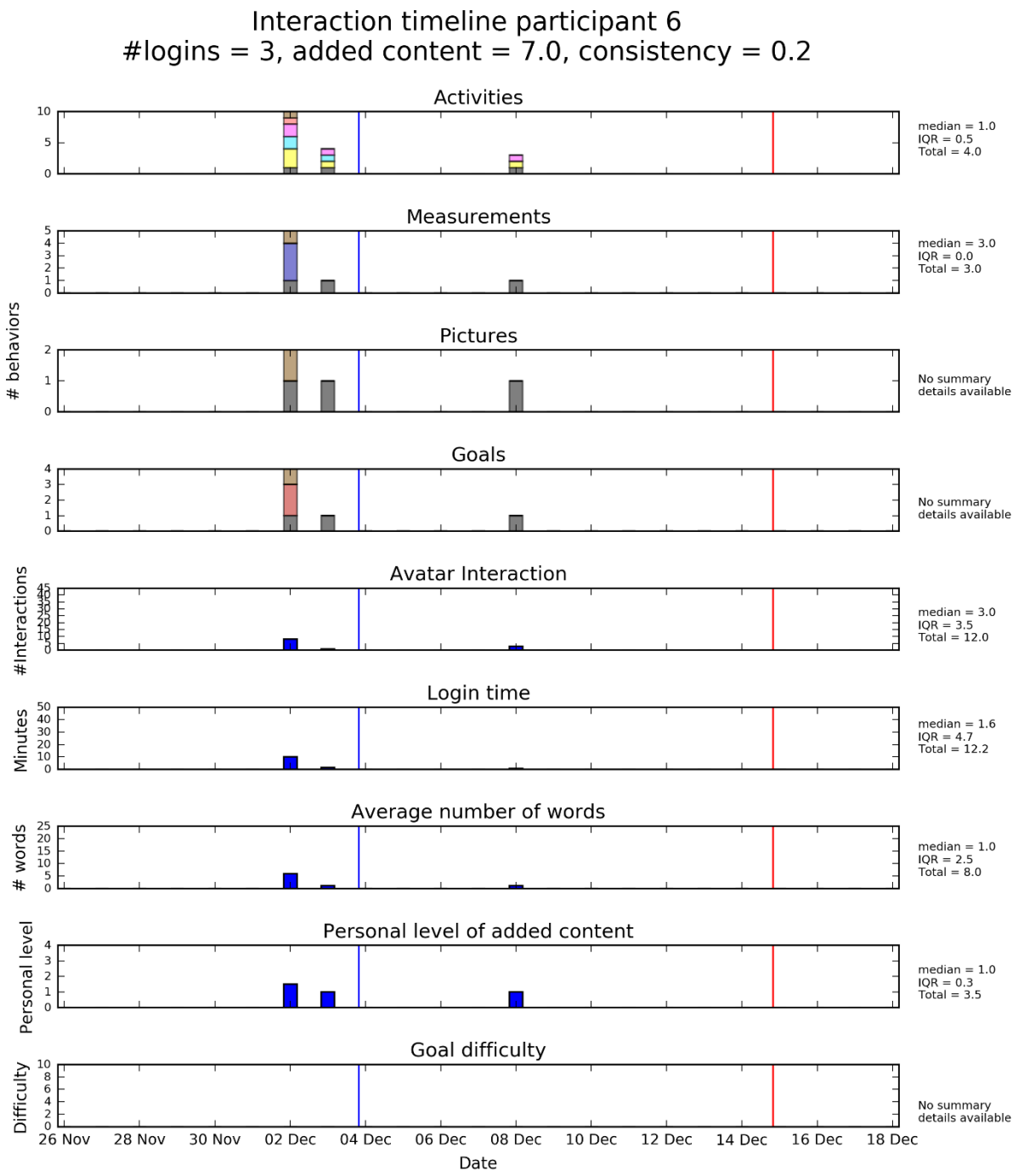


Figure 71: interaction timeline participant 6

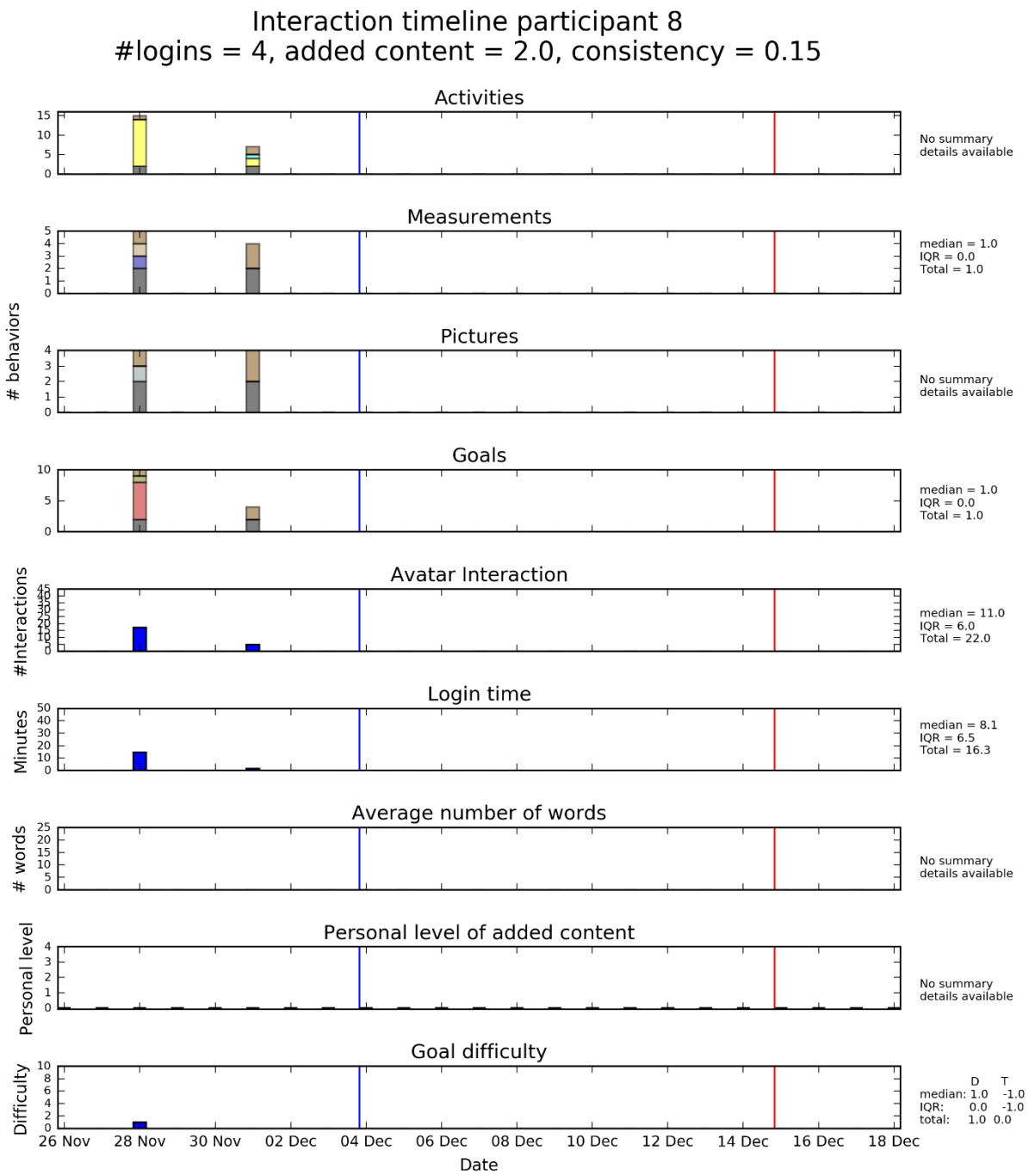


Figure 72: interaction timeline participant 8

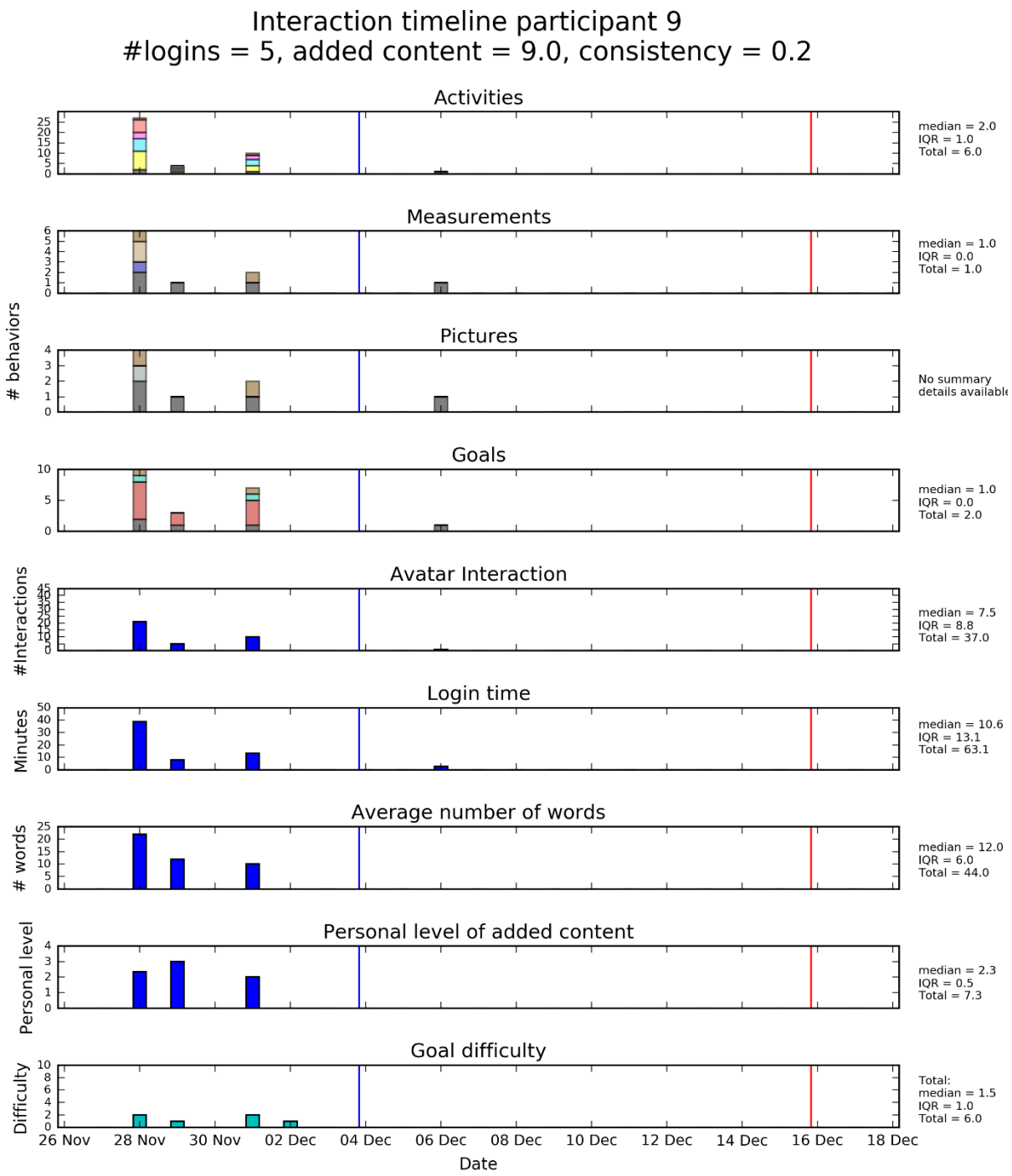


Figure 73: interaction timeline participant 9

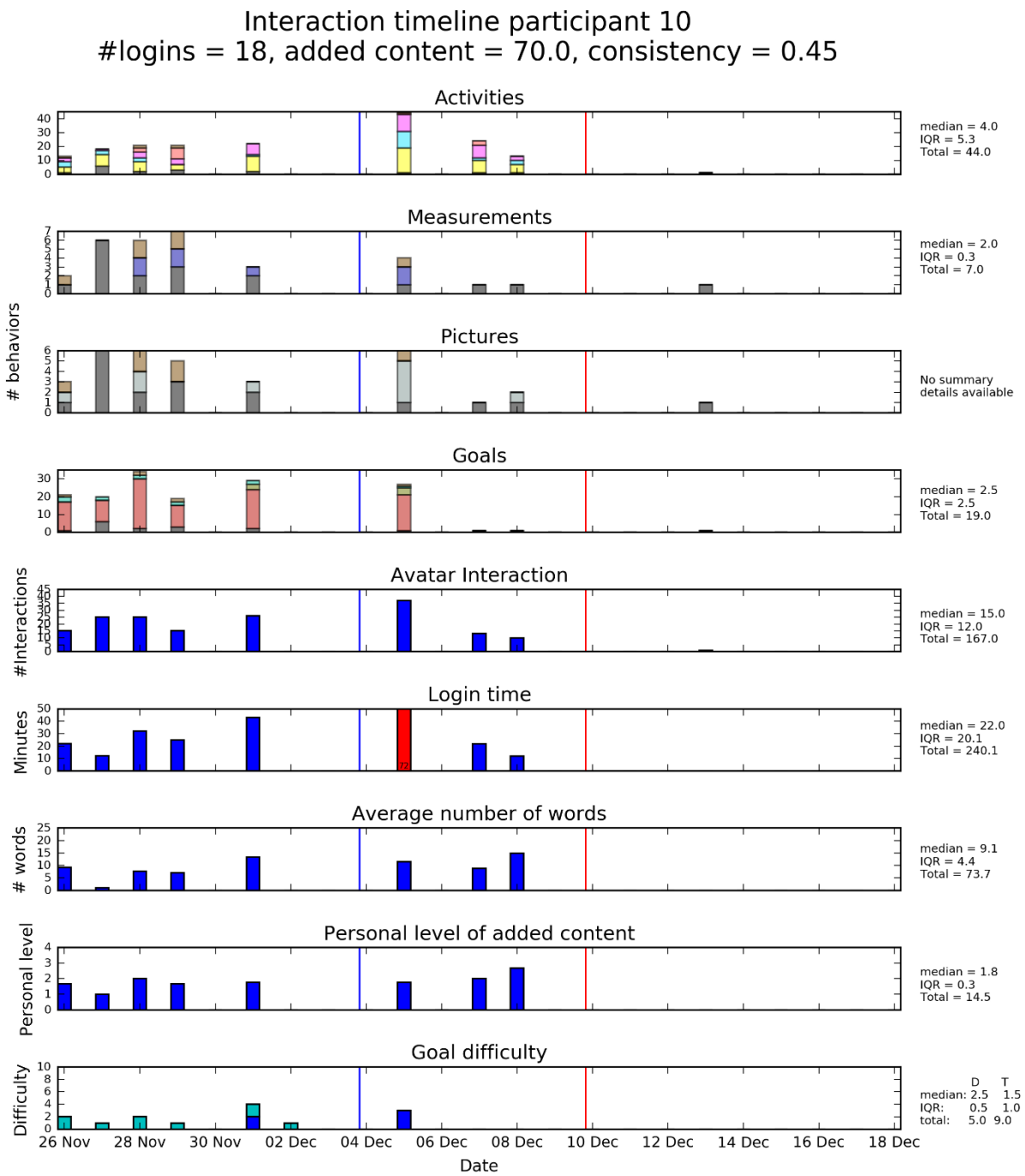


Figure 74: interaction participant 10

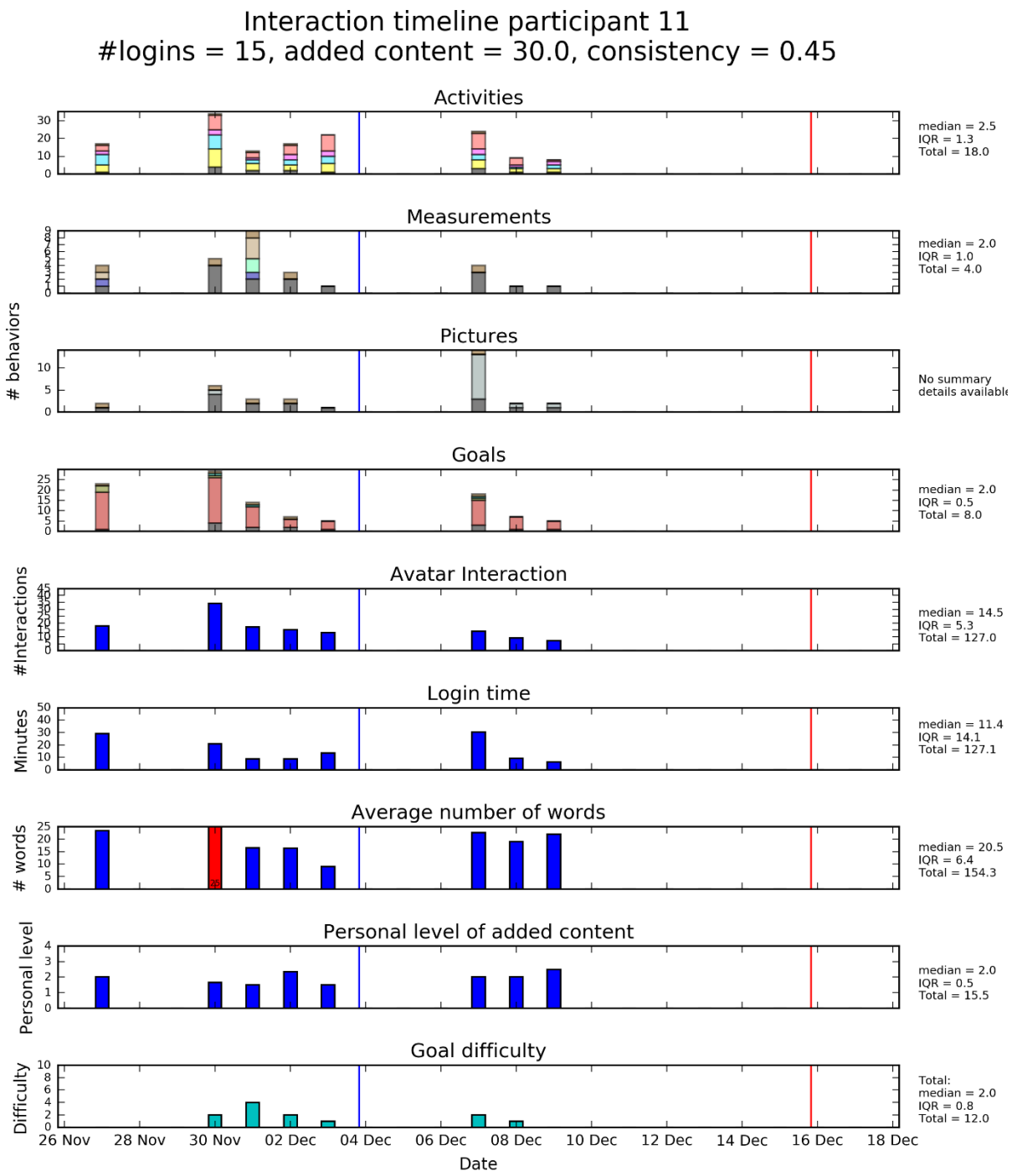


Figure 75: interaction timeline participant 11

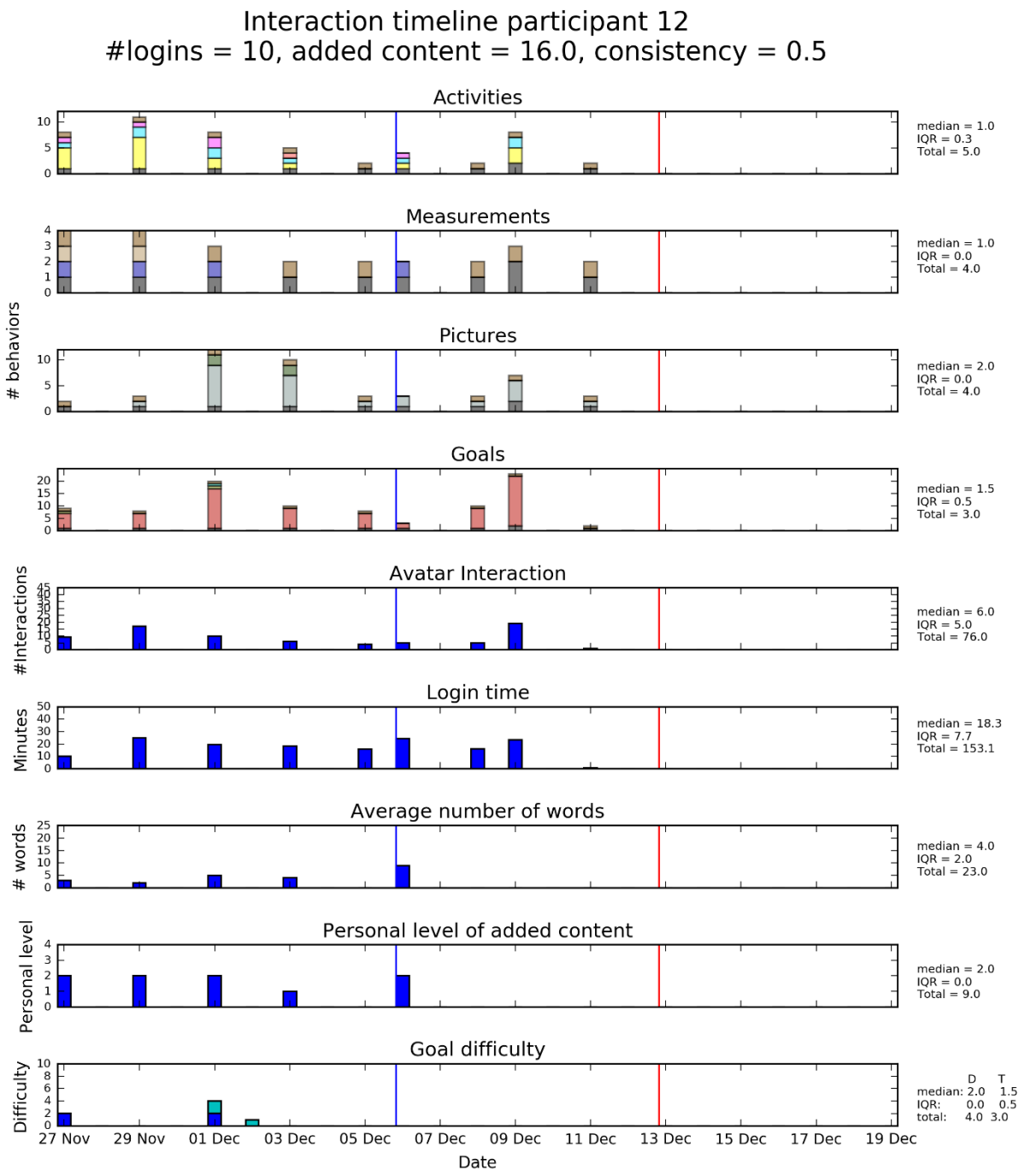


Figure 76: interaction timeline participant 12

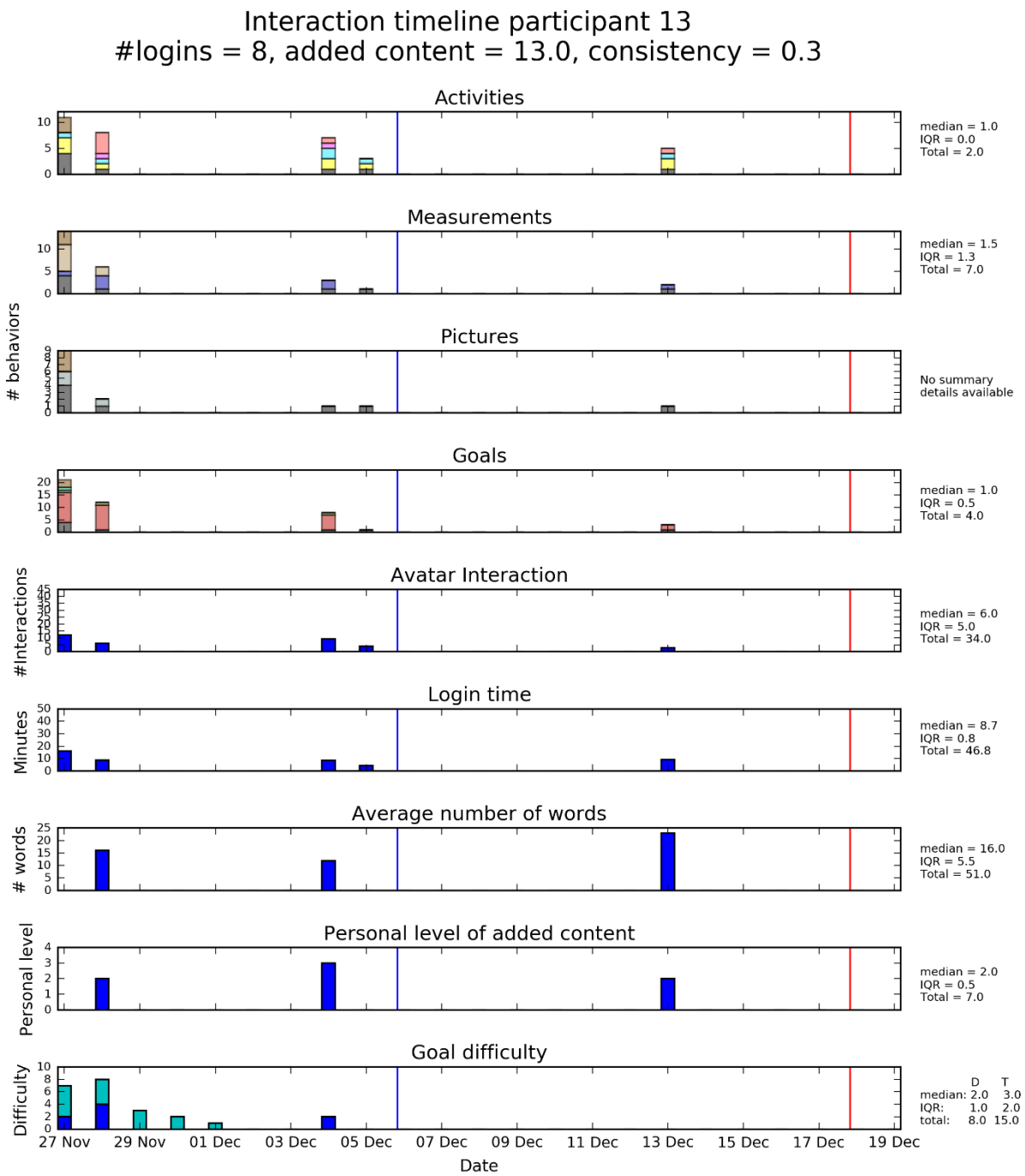


Figure 77: Interaction timeline participant 13

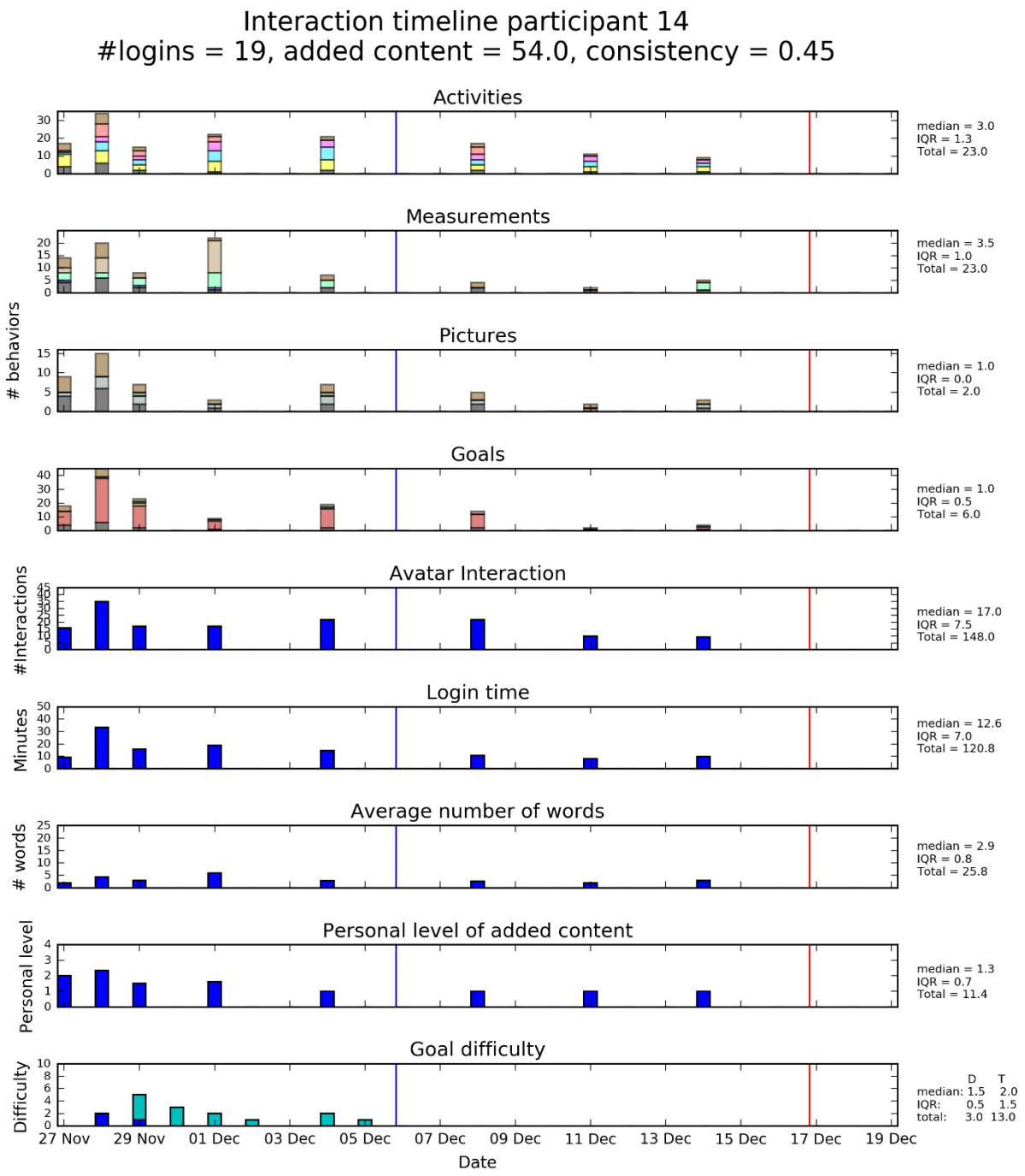


Figure 78: interaction timeline participant 14

4.2. Motivation timeline per user

For each user a timeline is constructed that displays the development of motivation and its antecedents feeling of autonomy, competence, and relatedness.

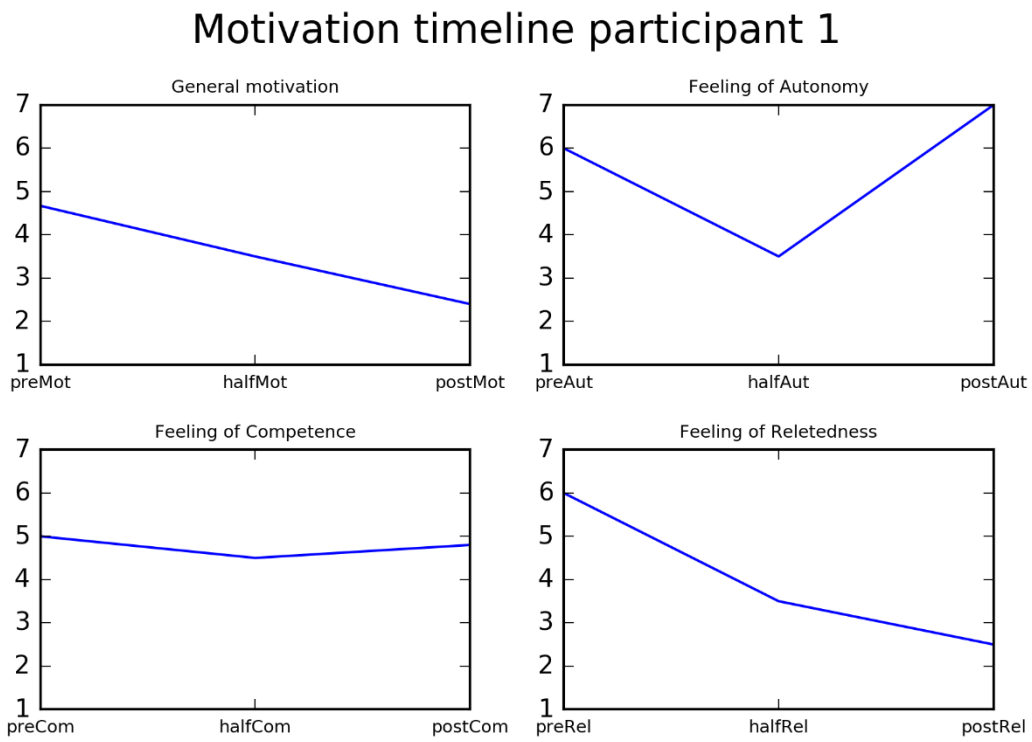


Figure 79: motivation timeline participant 1

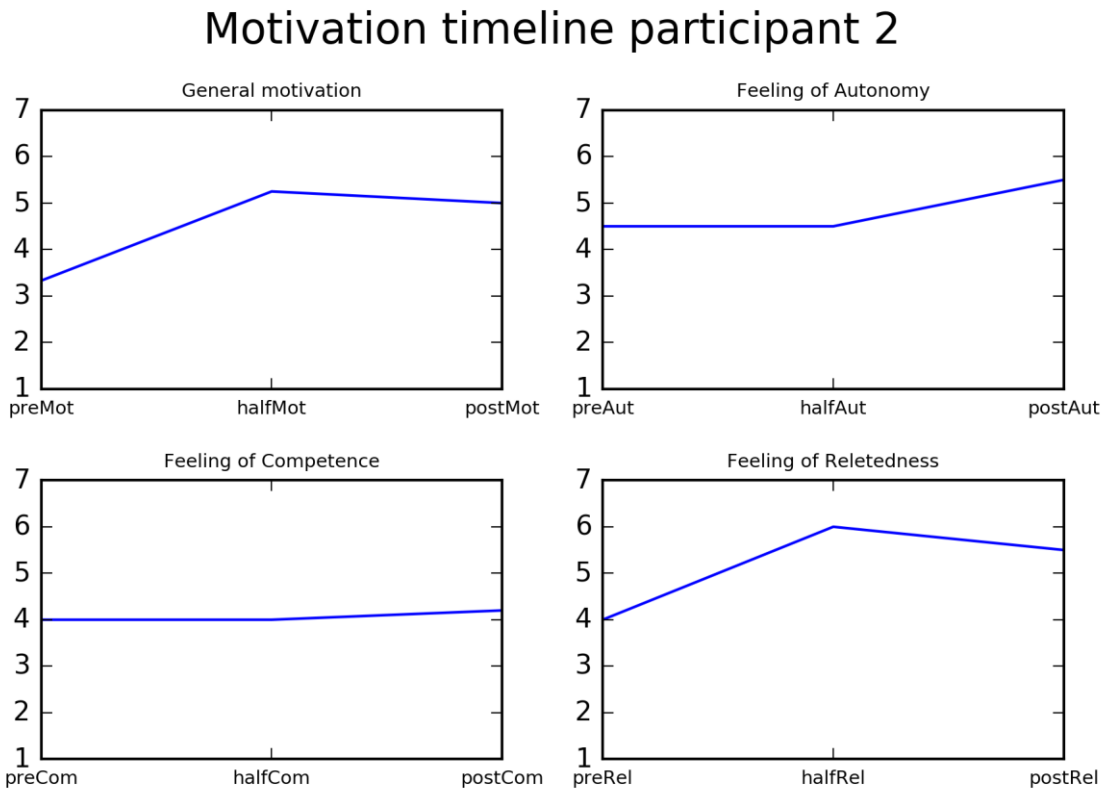


Figure 80: motivation timeline participant 2

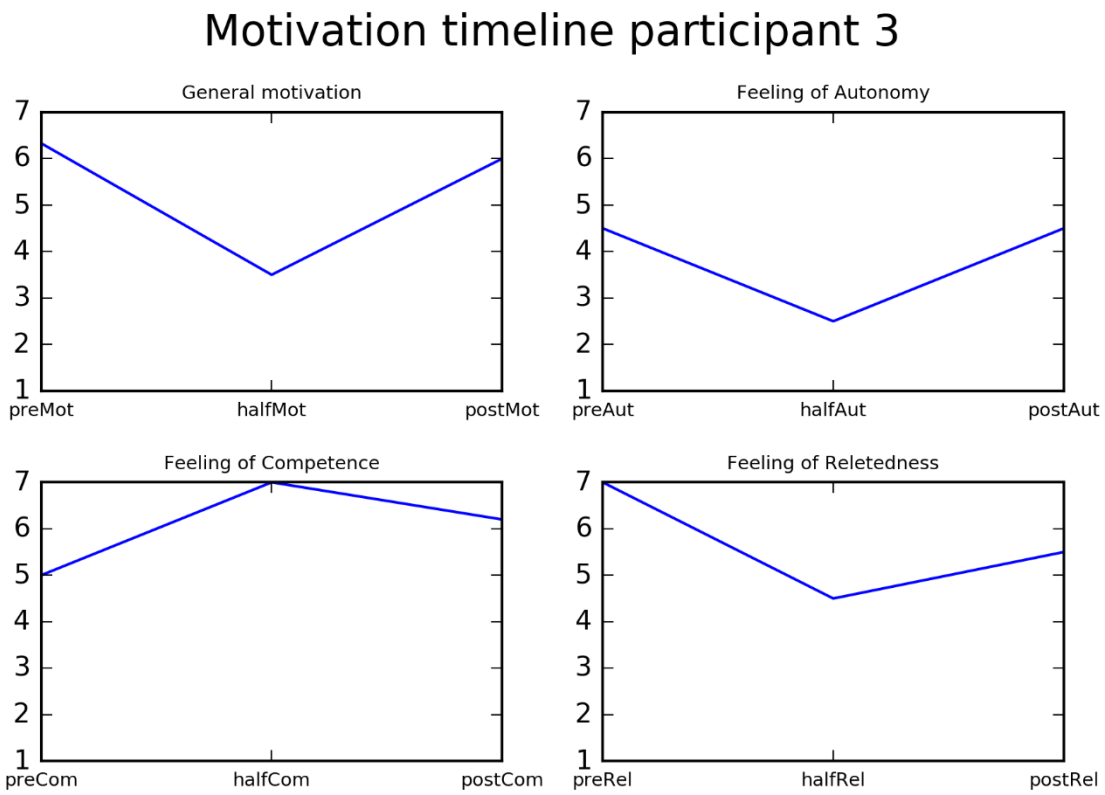


Figure 81: motivation timeline participant 3

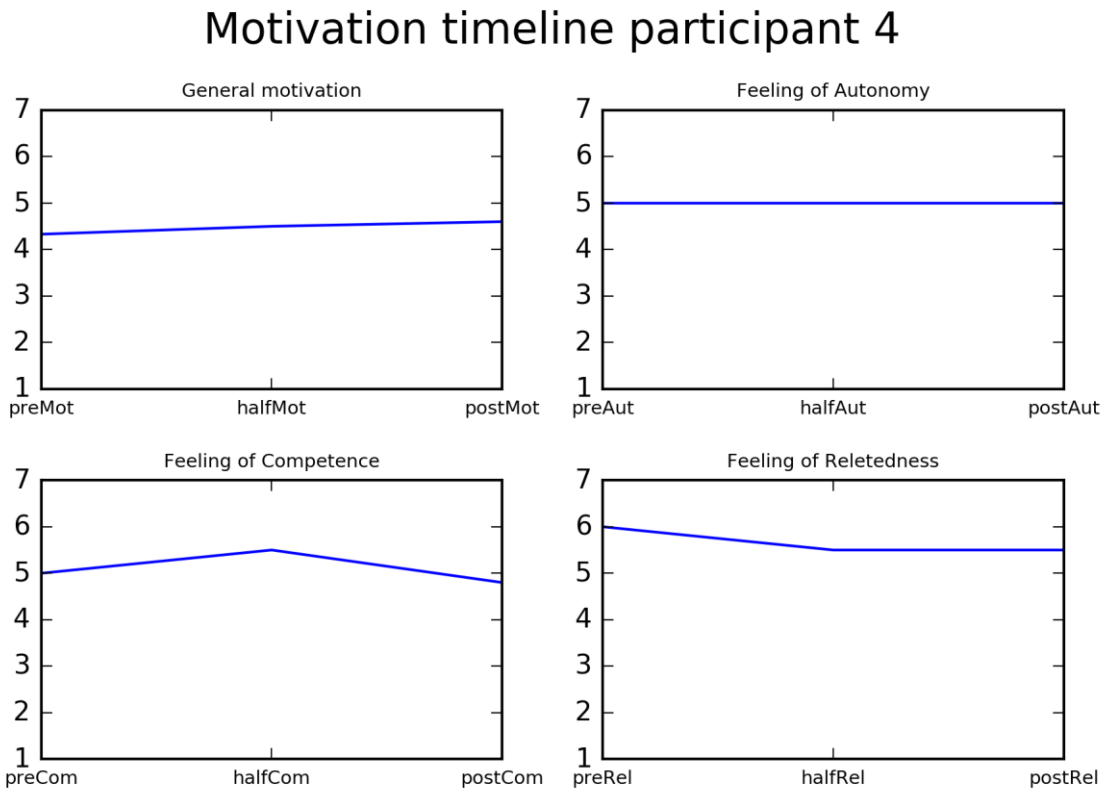


Figure 82: motivation timeline participant 4

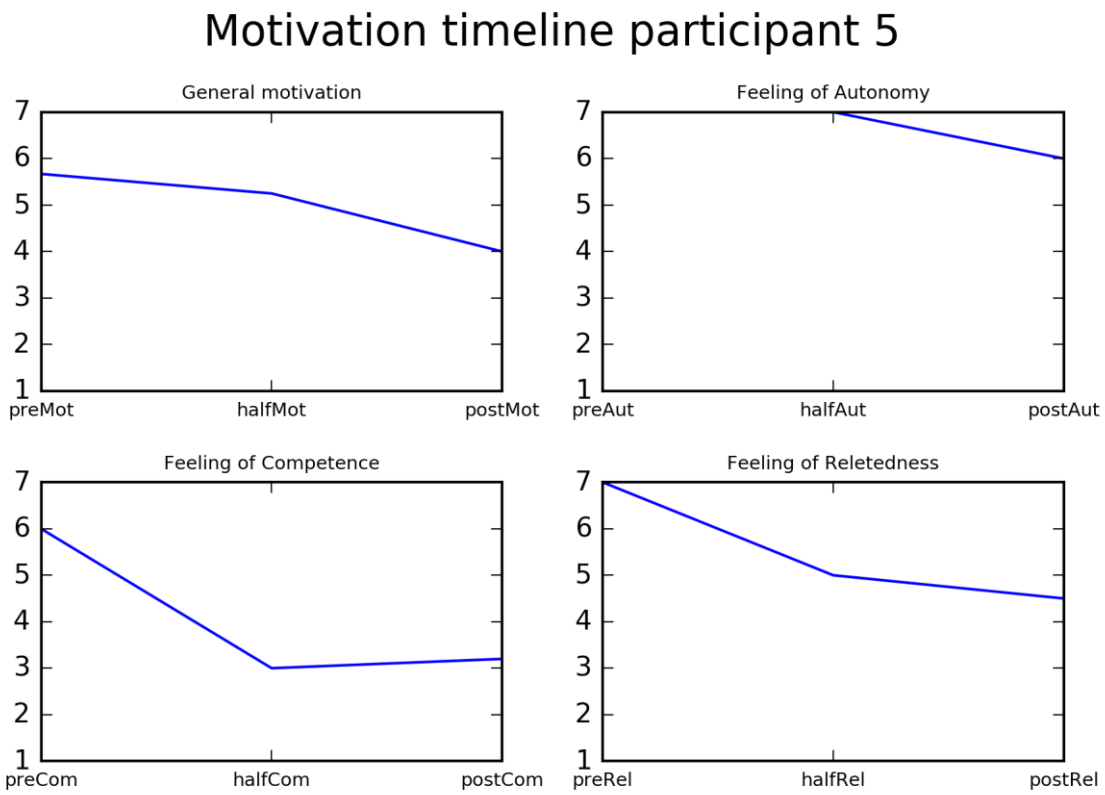


Figure 83: motivation timeline participant 5

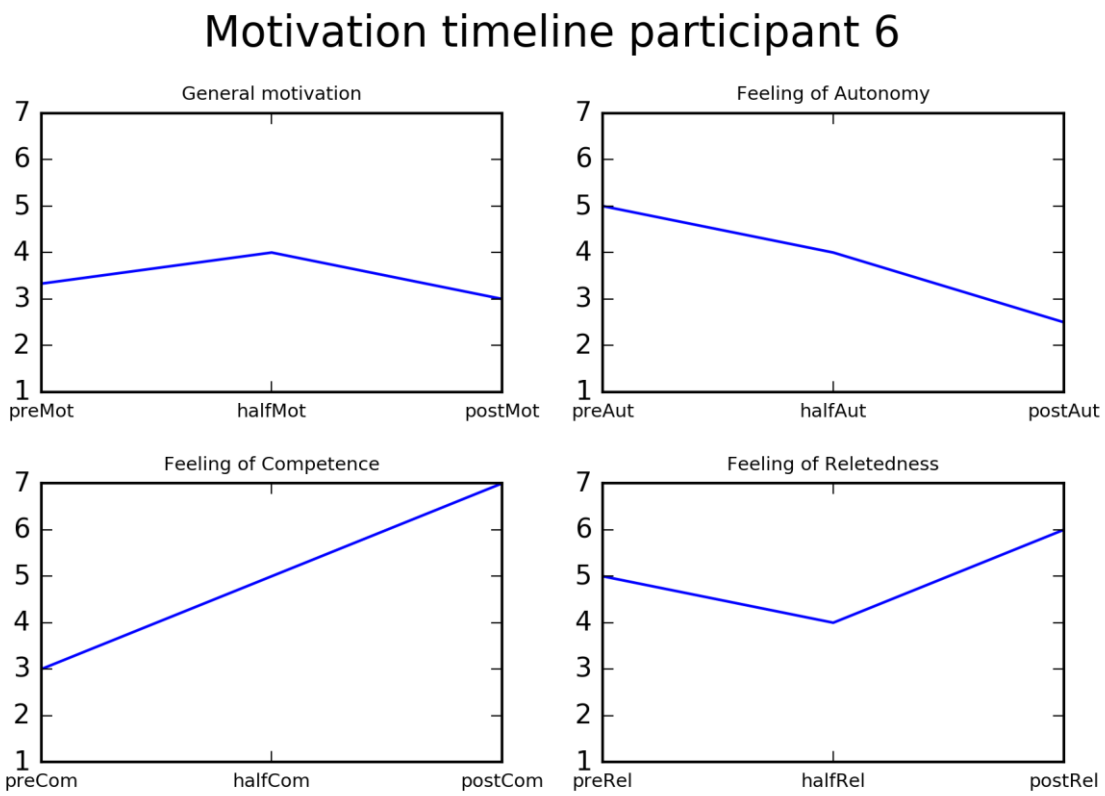


Figure 84: motivation timeline participant 6

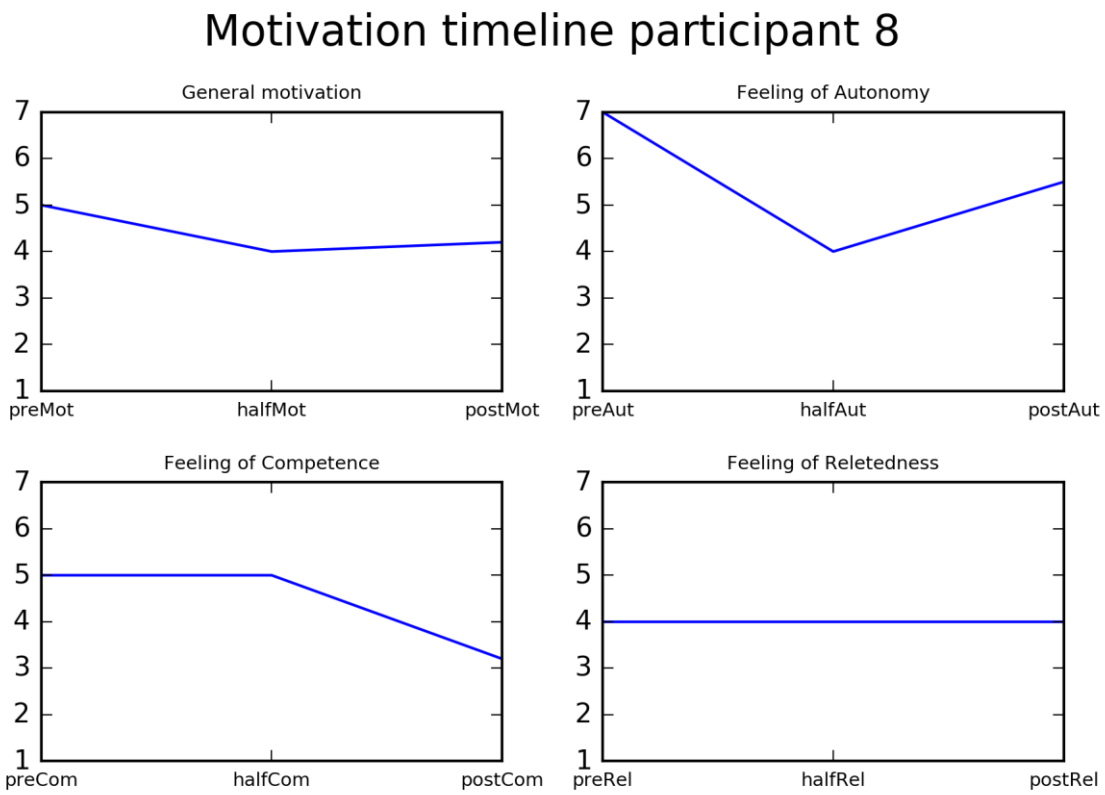


Figure 85: motivation timeline participant 8

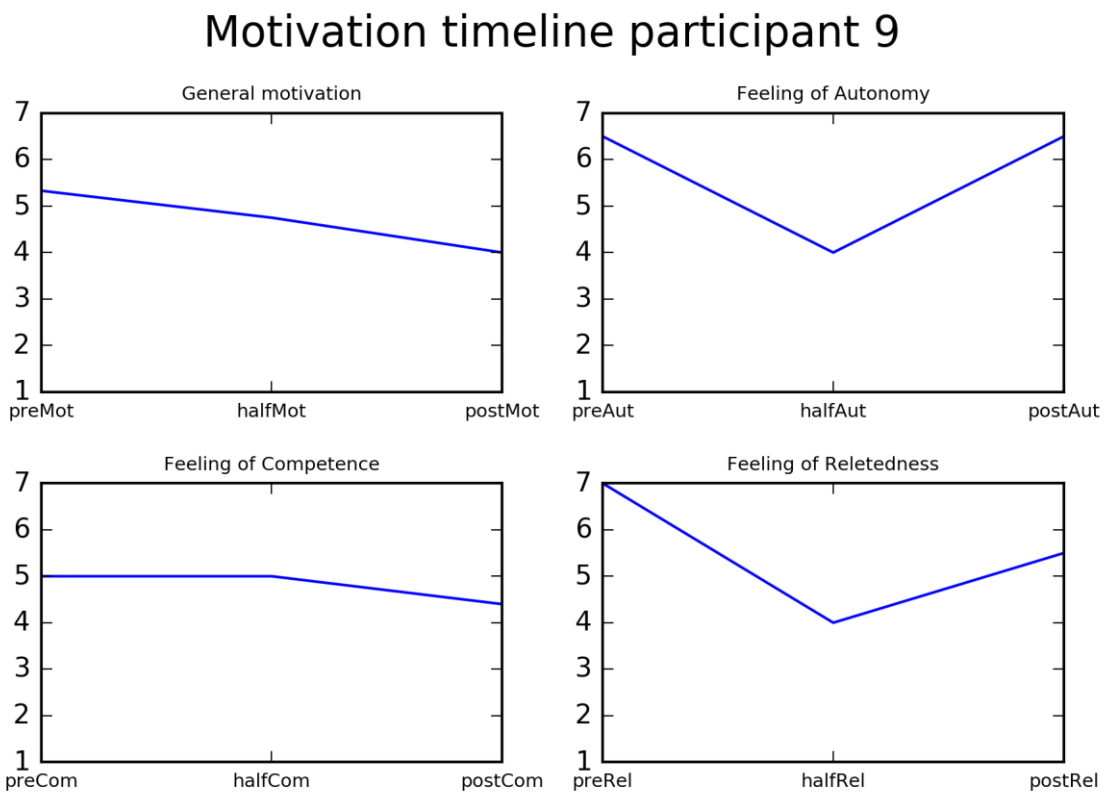


Figure 86: motivation timeline participant 9

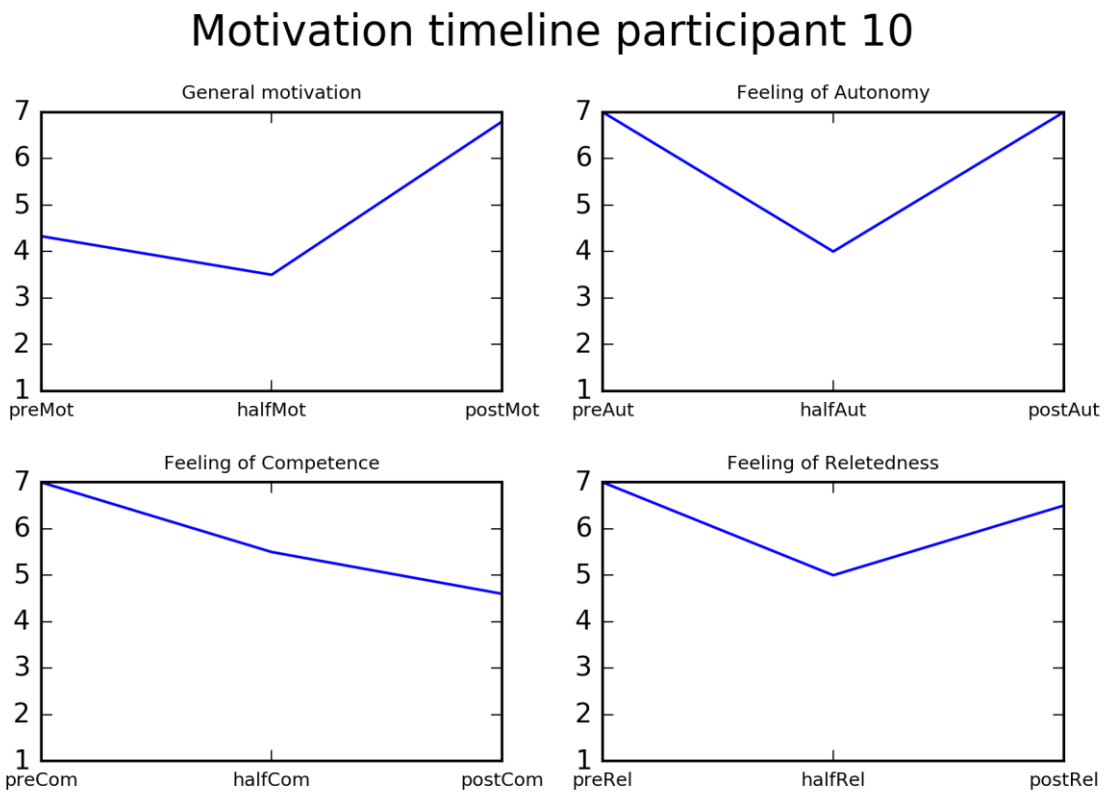


Figure 87: motivation timeline participant 10

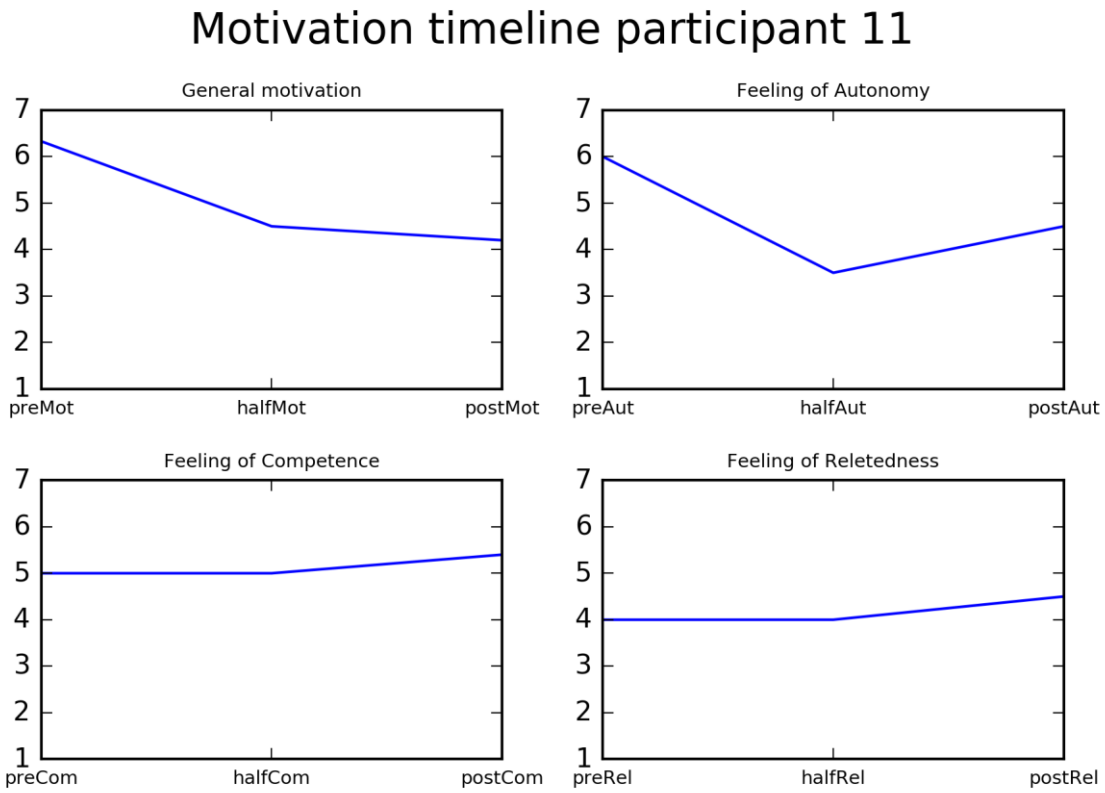


Figure 88: motivation timeline participant 11

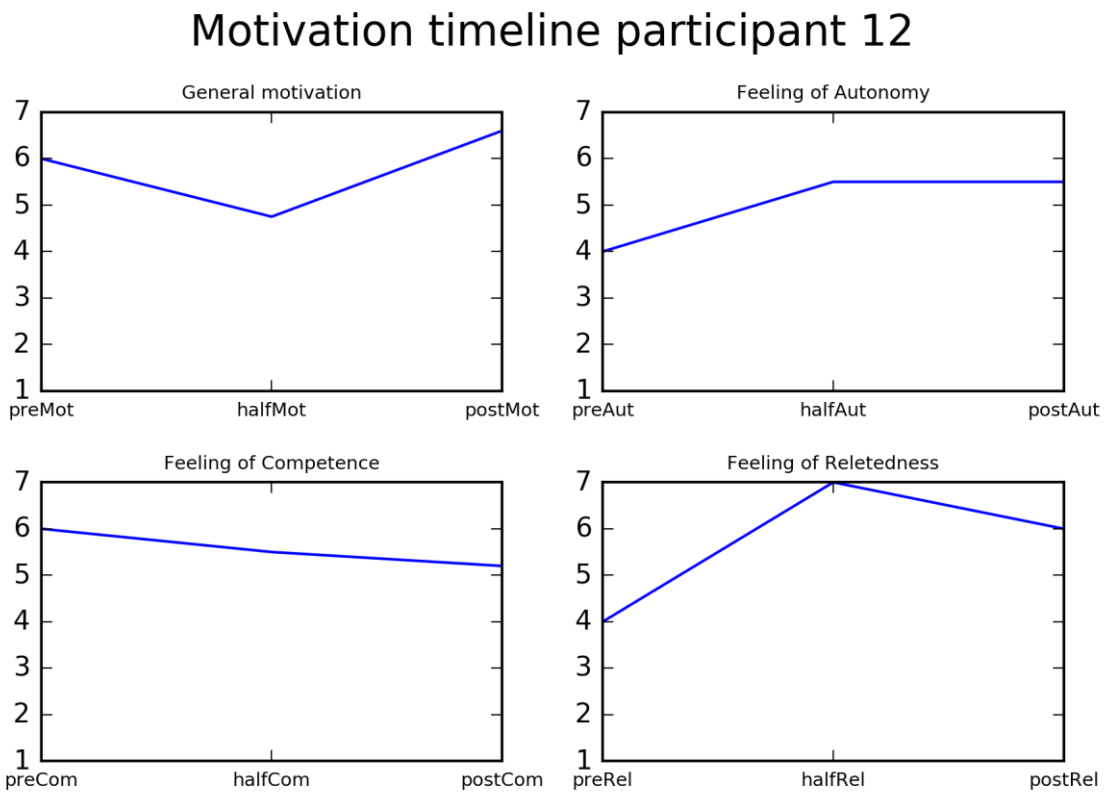


Figure 89: motivation timeline participant 12

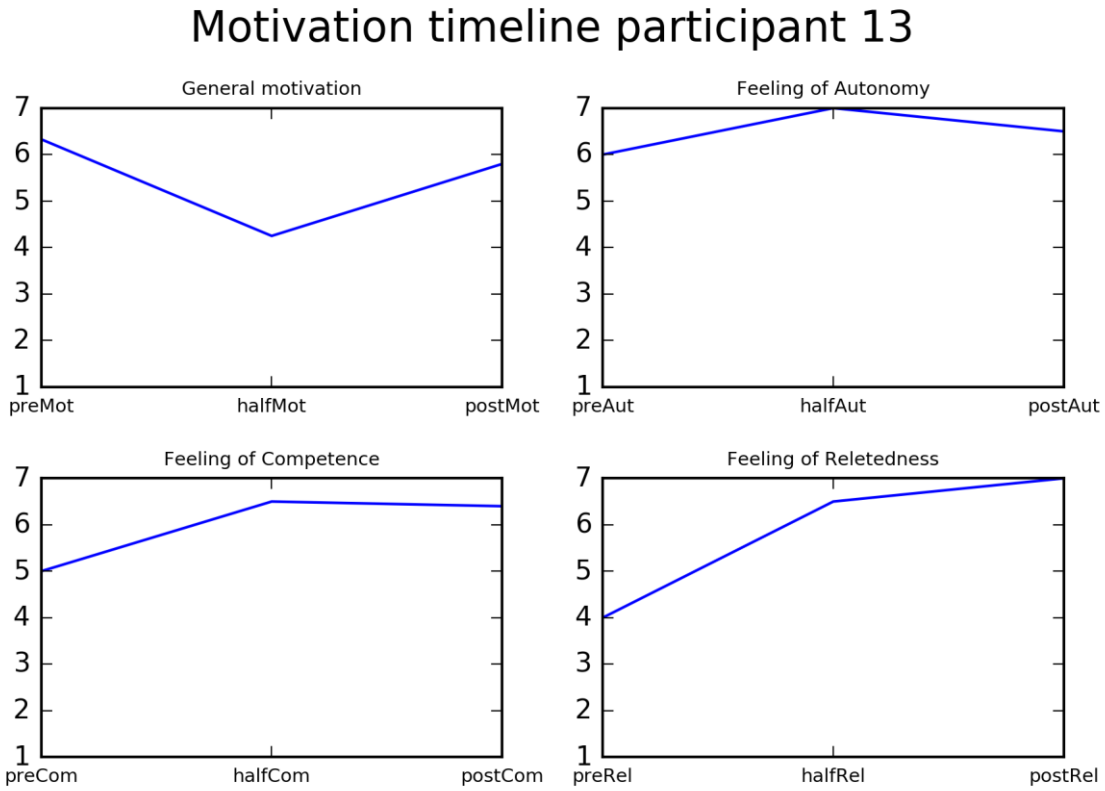


Figure 90: motivation timeline participant 13

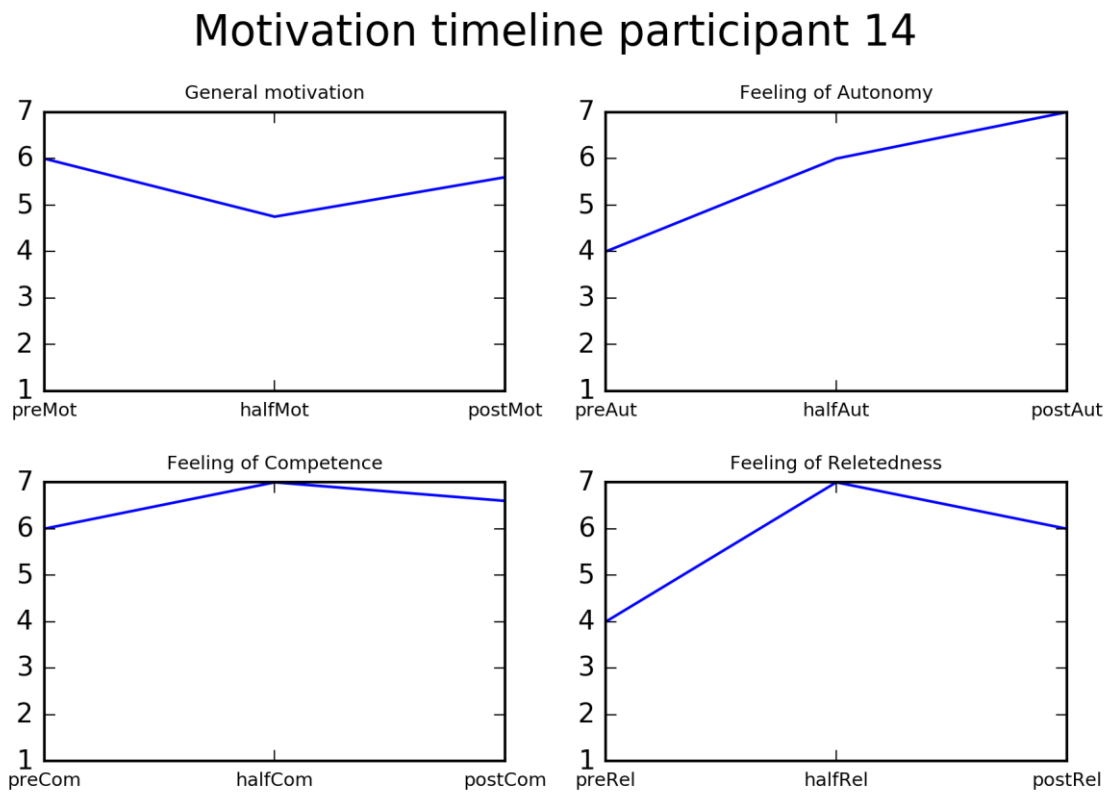


Figure 91: motivation timeline participant 13