User evaluations of a behaviour change support system

Master Thesis

Author:
Saskia M. Akkersdijk

Supervisors:
dr. E.M.A.G. van Dijk
(1st supervisor)
dr.ir. H.J.A. op den Akker
R. Klaassen, MSc

February 22, 2013

UNIVERSITY OF TWENTE.

Contents

1 Summary

In this thesis three versions of the SmarcoS-system are evaluated. SmarcoS-diabetic focuses on recently diagnosed diabetes type II patients. It gives feedback on medication intake and activity. Feedback is given in the form of text and can be received on a smartphone application and at a computer application. Secondly, two versions of the SmarcoS-office worker system are evaluated. These versions focus on office workers with an intense digital lifestyle. Both versions of the system only give feedback on activity. Feedback can only be received at a smartphone application, and is given in the form of text or by an Embodied Conversational Agent.

The implementation of the SmarcoS-diabetic system lead to participants manipulating the system to be registered as being on time to take their medication. They took their pills out of the pill dispenser before they took them to open the pill dispenser within the set time. Subsequently some participants forgot to take that medication. Introducing this kind of errors should be avoided.

In both evaluation we found that participants would like to omit the need to dock the activity monitor of the system. Now data is only known by the system after docking, while participants would like the system to have real-time data access. This also has implications for the feedback that can be given. Now feedback can only be given after the activity monitor was docked, making feedback most of the time too late.

Participants thought the content of the messages was standard, administrative, not motivating, not diverse enough and there were not enough different types of messages. Content can be improved by making it more specific, concrete, personal and preference based. These results are found in both evaluation, although more prominent in the SmarcoS-office worker evaluation.

We found a difference in the review of, and attitude towards the system between the two evaluations. This difference is best explained by the difference in target groups. The main difference between participants is that participants of the SmarcoS-diabetic evaluation are diabetics, while the participants of the SmarcoS-office worker evaluation are office workers. The main difference between these two groups is that diabetics have a disease and therefore can be considered patients. The office workers had no disease, and therefore could not be considered patients. This difference also makes it likely that the participants of the SmarcoS-diabetic evaluation are more conscious that they need to have a healthy lifestyle and take their medication on time. Therefore, they probably have a bigger interest or at least are more conscious of their interest in such behaviour change support system. It is therefore likely that they are more willing to give up privacy, because they get more advantages of giving up privacy compared to the non-patient office workers.

2 Introduction

We live in a world in which technology plays a big role. It helps us and serves us in accomplishing great things, but it can also help us live a more healthy and balanced life. We humans are creatures of habit, and when we want to break this sometimes we can use a little help. Technology can help us with that. It can analyse our behaviour, and personally motivate us without using 'expensive' human beings as coach. Technology can support the human care provider, and persuasive technology might play an important role in accomplishing the behaviour change, adherence to the new behaviour and the self-management role. A big advantage of persuasion by technology over human persuasion is that technology is more persistent and can go where humans cannot. Personalised feedback can be given based on individual performance in relation to the goal. In this thesis we will look at technology that can help us change our behaviours. We will be looking at and evaluating a behaviour change support system. This system is the SmarcoS system. We will now briefly tell something about the SmarcoS-project and the system.

SmarcoS and Attentive Personal Systems SmarcoS is a European project that aims to help users of interconnected embedded systems by ensuring their inter-usability. SmarcoS allows 'devices and services to communicate in user interface (UI) level terms and symbols, exchange context information, user actions, and semantic data. It allows applications to follow the user's actions, predict needs and react appropriately to unexpected actions. The use cases for the project are constructed around three complementary domains: attentive personal systems, interusable devices and complex systems control [3]'. This research focuses on part of the attentive personal system.

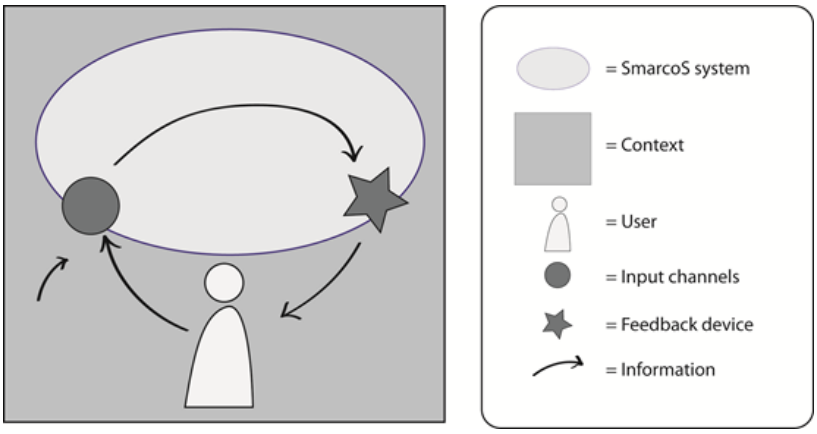


Figure 2.1: Graphical representation of basic elements for SmarcoS feedback models

One of the goals of the SmarcoS project is to create an intelligent system that motivates and supports consumers in their daily life to live a balanced and healthy lifestyle using the notion of inter-usability and task-driven UI modelling technologies. Part of this work package is an attentive personal system that targets healthy consumers as well as chronic patients. For healthy consumers this system aims to support them to live a healthy and balanced lifestyle. For chronic patients the system aims to reduce medical complications by better managing their condition through a combination of

self-monitoring, education and qualitative analysis, while reducing costs for care givers, employers and insurance providers.

There are two use cases in which the system should operate. The target group in the first use case consists of office workers with an intense digital lifestyle; the system should encourage them to live a more active life and make healthy choices. In the second use case the target group consists of recently diagnosed diabetes type II patients. The system should help them monitoring their glucose levels, medication intake, activity level and making the healthy choice. Based on the written scenarios of each use case (five for office workers, two for diabetes type II patients) the consequences for attentive personal systems and how they should support the feedback models were investigated. These consequences and feedback models form functional and non-functional requirements that are taken into account when designing the attentive personal system.

An important part of the attentive personal system is providing feedback at the right time, while taking the context of the receiver of that feedback into account. The feedback model identified for this system defines the interactions with a user in a given context. The basic elements for SmarcoS feedback models are shown in figure 2.1. We see that information about the context is given to the system, via the input channels. There the information is processed and output is generated. This output, in the form of feedback, is then given by the system to the user via the feedback device.

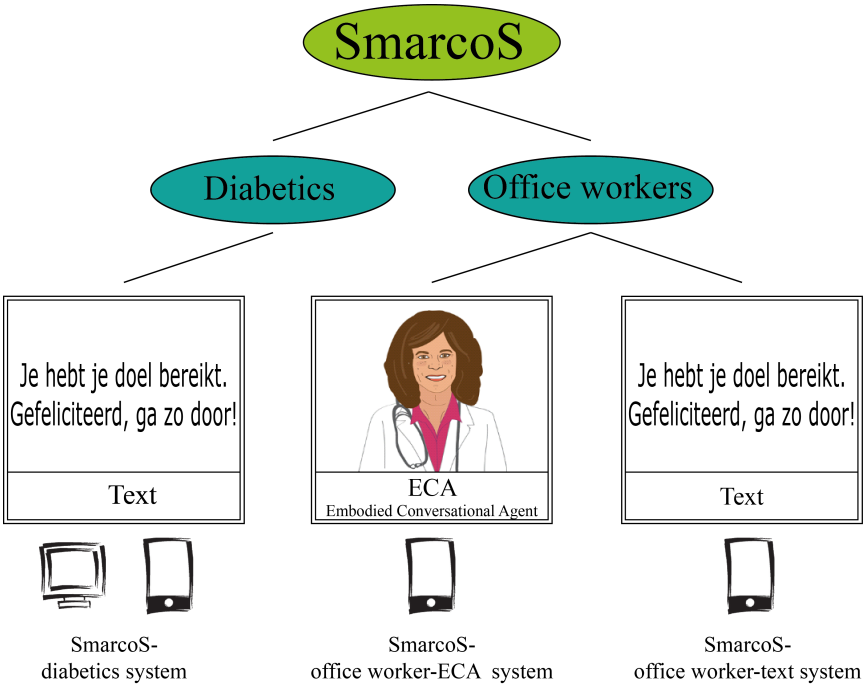


Figure 2.2: Basic representation of the different versions of the system. The SmarcoS system consists of three different versions, one for each target group. And for the office worker target group a second version was made with a different modality (ECA). All versions had a smartphone application, while the SmarcoS-diabetics group also had a computer application.

Versions of the system In this thesis we will look at three versions of the SmarcoS system. To accommodate all requirements for each target group two text based versions were made (one for each target group). Both versions give feedback on activity level, while the diabetic version also accommodates medication intake and reminders.

Because we also wanted to investigate modality a third version was made, with an embodied conversational agent. That version was used with office workers. The only difference between the text version and the ECA version for office workers is the modality. These versions and their target group are shown in figure 2.2. All versions share the same basis (see section 5.1), but use a different set of rule to generate feedback. The SmarcoS-diabetic system has two application (one smartphone application and a computer application) via which the user can receive feedback. The SmarcoS-office worker system only has one application via which the user can receive feedback. We will discuss the SmarcoS-diabetic (SD) system in section 5.1.2, while we will discuss the SmarcoS-office worker systems (SO) (SmarcoS-office worker-ECA (SOECA) and SmarcoS-office worker-text (SOT) in section 6.1.

In this thesis we will evaluate all versions of the system. We will present the results, discuss them and state the conclusions we can make from these results. Finally, we will look at future work.

In the next chapter we will look at theory of behaviour change support systems, and embodied conversational agents. Of behaviour change support systems we will discuss two psychological theories (Goal-Setting Theory and Transtheoretical model of Behaviour Change). Secondly, we will discuss the different generations of behaviour change support systems. Finally, we will look at the elements of behaviour change support systems and discuss several important concepts. In chapter three we will give an overview of possible evaluation questionnaires that we looked at for evaluation the SmarcoS system. We will look at questionnaires measuring, usability, user interface satisfaction, user experience, technology acceptance, robot acceptance, source credibility and coaching's quality. In chapter four we will further discuss the shared basis of the SmarcoS system, and explain the version for diabetics (SmarcoS-diabetic). In this chapter we will also give the methodology and results of the evaluation of that version of the SmarcoS system. Finally we will discuss the results of this evaluation. We will do the same for the other two versions (SmarcoS-office worker) in chapter five. The discussion of the differences between the versions of the system and a reflection to the theory can be found in chapter six. Chapter seven will contain the conclusions. We will end this thesis in chapter eight by looking at future work.

3 Theory

In this chapter we will first look at behaviour change support systems in general. We will look at important psychological theories, the various generations of behaviour change support systems we can distinguish, and elements of a behaviour change support system. Secondly, we will discuss Embodied Conversational Agents. We will tell what an Embodied Conversational Agent is, discuss some of the existing agents, and discuss the ECA used in the SmarcoS system.

3.1 Behaviour change support systems

The main goal of the SmarcoS system is to motivate and support consumers in their daily life to live a balanced and healthy life. The system therefore encourages behaviour change from a less healthy lifestyle. In this section we will look into behaviour change support systems. We will look at two of the psychological theories on which these systems are built, look at different generations of behaviour change support systems and discuss elements of these systems that are important when building a behaviour change support system.

3.1.1 Psychological theories

Behaviour and how to influence behaviour is studied in the field of psychology. Systems meant to change behaviour or that should persuade people to do something are often based on behavioural theories such as Goal-Setting Theory or Transtheoretical model of Behaviour Change [20]. We will now shortly discuss both theories, starting with the Goal-Setting Theory.

Goal-Setting Theory The Goal-Setting Theory states that there is a relationship between the difficulty and specificity of a goal and the performance of the task. Specifically, difficult goals consistently lead to higher performance than urging people to do their best.

When people are asked to do their best they do not do so, because these goals have no external reference. Goals with no external reference allow for a wide range of acceptable performance levels, in contrast with specific goals. However, having a specific goal does not automatically mean that people perform better, because specific goals vary in difficulty. However, having a specific goal does help to reduce the ambiguity of what is to be attained [60].

Having a goal affects performance through four mechanisms. First, a goal helps to direct attention and effort toward activities that are goal-relevant, while directing attention away from goal irrelevant activities [61]. Secondly, high goals lead to a greater effort than low goals. Therefore, goals have an energizing function [61]. Thirdly, goals help with persistence. Hard goals prolong the effort of people [56]. Finally, goals also affect actions indirectly, because they can lead to arousal, discovery, and talk about relevant knowledge and strategies [61]

There are three important moderators of goal effects. First of all, goal commitment enhances the relation between a set goal and the performance. Commitment is most important and relevant when goals are difficult [52], because difficult goals are associated with lower chances of success and require more effort [28]. Goal commitment is facilitated by the importance of the goal and by self-efficacy. When a goal is more important to

someone, he/she will be more committed to it [61]. And when people believe that they can attain a goal, self-efficacy is increased which enhances goal commitment [61]. Secondly, appropriate feedback that reveals progress in relation to the goal is important, because it creates a possibility for people to adjust the level or direction of their effort or to adjust their performance strategies to match what the goal requires. The combination of goals with feedback is more effective than goals alone [5, 27]. Finally, task complexity is a moderator of goal effects. If you have complex tasks, higher level skills and better strategies are needed. These need to be attained, and this takes time [61].

Transtheoretical model of Behaviour Change The Transtheoretical model of Behaviour Change describes the process people go through when changing their behaviour. This model includes stages of change to integrate processes and principles of change from different theories of intervention. Change is seen as a process involving progress through a series of six stages.

The first stage is precontemplation. In this stage people do not intend to take action in the foreseeable future (the next six months). In the second stage, contemplation, people become more aware of the pros of changing, while they are also acutely aware of the cons. This is the stage in which people intend to change something in the next six months. However, people also can keep stuck in this stage because their pros and cons balance. In the third stage, preparation, people have an action plan. They intend to take action in the immediate future (the next month). Action is the next stage, in which people actually make modifications to their life. After this the maintenance stage is reached, in which people try to prevent relapse. This stage lasts from six months to about five years. Termination is the final stage. During this stage people will not return to old habits, no matter what is happening to them. The activities that people use to progress through the stages we call processes of change. Ten processes are identified being; consciousness raising, dramatic relief, self-re-evaluation, environmental re-evaluation, self-liberation, social liberation, counterconditioning, stimulus control, contingency management and helping relationships [79, 81, 80]. We will not explain these processes in detail, and refer the interested reader to the mentioned articles.

Psychological theories and their relation to the SmarcoS system The discussed psychological theories are important because they give a background on which the system builds. The SmarcoS system mostly uses the Goal-Setting Theory. The system helps users to set realistic goals, and tries to motivate the users to be more active by giving feedback on the activity level of that moment. When we look at the stages of the Transtheoretical model of Behaviour Change we can place the users of the SmarcoS system on the fourth stage; the action stage. As discussed above, action is the stage in which people actually make modifications to their life. The system helps to make these modifications. When users are not in the correct stage of behaviour change, the system is less likely to improve the activity level of users.

3.1.2 Generations of behaviour change support systems

When we look at behaviour change support systems eHealth technologies allow for more individualized behaviour change interventions. EHealth can be seen as the use of emerging information and communication technology, especially the Internet, to improve or enable health and health care. We can distinguish between several generations of this kind of behaviour change support systems to up to three generations.

The first generation of these system facilitated intervention tailoring with computers to generate printed materials. Examples of this kind of material are pamphlets, newsletters, reports, and magazines [16, 68, 66].

Second generation interventions are delivered through interactive technology or desktop applications such as websites, email and CD-ROM programs [46, 68, 15]. This second generation allows for direct interaction between the participant and the technology, this increases capabilities beyond tailored feedback messages. It also can give participants access to educational information, report on goal and track their progress. But also allows social support via bulletin boards, or synchronous chat rooms [68].

Third generation technologies include mobile devices such as handheld computers, cell phones, and text messaging devices. This enhances the potential for timely feedback and assessment [68]. New functions can be incorporated such as, sensing, monitoring, geospatial tracking, and location-based knowledge presentation [68, 75]. This also enhances possibilities for accurate assessment and tailored feedback.

In section 5.1, and 6.1 we will discuss which generation each version of the SmarcoS system belongs to.

3.1.3 Elements of a behaviour change support system

When building a behaviour change support system it should of course be a useful system. A useless system will not be used, no matter how easy it is, or how nice it looks [69].

As we already saw in the Goal-Setting Theory, setting specific goals is important when behaviour change is wanted. Behaviour change support systems can help to set those specific goals, and help people change behaviour by reminding them of their goals. As we already saw, the combination of a specific goal plus feedback about progress towards this goal is more effective than goals alone [92, 61, 62, 58]. Therefore, a system that is capable of given feedback about progress, apart from helping to set goals will be more effective.

It is also important to give information about the behaviour change to the user; why is the behaviour change important? What are the benefits of the behaviour change? How does the system help you change your behaviour? Providing information to the user allows for making informed decisions.[30, 58].

Informing users and help them set their goals, while giving feedback, helps to reduce any barriers that people experience when going through a behaviour change. This increases the likelihood of certain behaviour and makes people more confident about the behaviour change [30, 4, 44]. It also makes the goal behaviour seem more achievable. When behaviour seems more achievable, self-efficacy of people is increased [30, 4, 59]. Behaviour change support systems can help to shape a person's mental model by channelling behaviour in a certain pattern [30].

To help creating a successful behaviour change support system several concepts are of importance. These concepts can help in different ways. First of all, if the system has an authority role, such as a counsellor or an expert, the system automatically gains the influence that comes with being in a position of authority. This influences the expectance of people; they expect the system to be intelligent and powerful. Praise from a system generates the same positive effects as praise from other humans [31, 32]. Secondly, when the message needs to be truly persuasive, it should be personalised to the user's interests and characteristics [53]. A message that is tailored to an individual is more effective than generic communication. Thirdly, when the system has a character that looks like a person that communicates this message people are more likely to cooperate with the message than when it is communicated by a clearly unreal computer character, even if they find this character appealing and likable [74]. Fourthly, when your system can be easily accessed on several devices, it will be more effective. A behaviour change support system that is able to intervene with several contact points is expected to be more effective in stimulating behaviour changes than those that use a single contact point [67, 74]. Fifthly, a more attractive technology will have greater persuasive power than an unattractive technology, and the mere appearance of a system is sufficient to

change its social influence [31, 74]. Finally, when a system is easy to use it increases the likelihood that the system will be used. A system that is really useful and attractive, but is difficult to use will be used by less people than a system that is easy [69].

To end this section we would like to give some attention to the ethical side of behaviour change support systems. This side should be treated with care, since these systems influence people. They can form, alter or reinforce attitudes, behaviours or an act of complying. What are 'good' reasons for having a system that tries to change behaviour? How much can you influence people without informing them about it? All behaviour change support systems should help behaviour change, but do this while avoiding deception, coercion or inducements [69, 30]. They should respect individual privacy and enhance personal freedom[30].

By designing a transparent system, while considering the above issues, trust is built towards the system, this also increases the chance of behaviour change [69].

We have now seen on which psychological theories behaviour changing support systems are built. We saw that there are several generations in behaviour change support systems, we discussed important elements of behaviour changing support systems and looked at other factors that influence the way these systems work. This functions as a background in which we can place the SmarcoS system.

3.2 Embodied Conversational Agents

An Embodied Conversational Agent (ECA) is a computer character with human-like behaviour, with or without human-like appearance. What distinguishes ECA's from other computer generated characters is that they display interactive behaviours. Most ECA's are designed to carry out face-to-face conversations with users. In these conversations appropriate use of conversational non-verbal behaviour is included, for example hand gestures and facial expressions [83, 18].

There are many different applications for ECA's. ECA's can serve as guides, receptionists, teaching agents, entertainment agents, and support agents. When ECA's are used for behaviour change, they are support agents. Examples of support agents are the Virtual therapist, Rea (a real estate agent), Psychometer (a virtual therapist), and Laura, the Bickmore agent (improving attitude towards exercising). We will shortly tell about each of these agents. After this we will discuss the ECA that is used in the SmarcoS system.

The virtual therapist is created by Pontier and Siddiqui [77]. They added a virtual head to an online self-report questionnaire. The virtual head supports users while filling it out. All 21 multiple-choice questions of this questionnaire are asked by the virtual character, and the character shows affective behaviour. The agent shows sadness if the answers of the user show depression, while the agent shows happiness if the answers show that the user is fine.

Rea, is a more complex virtual character, that was created by Cassell et al. [17]. She is a virtual real estate agent that is capable of showing the user a (virtual) house (see figure 3.1). Rea has a human-like body and uses her body during conversations. She uses eye gaze, hand gestures, body postures and facial expressions in the conversation. She is also capable of understanding some user input, being designed to respond to visual, audio and speech clues. While showing the rooms of the house, she provides information about the rooms and asks the user questions.

The Psychometer is more like the virtual therapist. The agent asks a series of five point Likert-scale questions to determine the 'personality' of the user in terms of five personality traits. The user can answer each question in a normal utterance, and the

agent tries to determine the exact answer. When the answer is not clear it will ask for clarification. The agent asks a set of questions one by one, and if the user asks for the meaning of a word it will provide this meaning [86, 84]. An image of the Psychometer with the agent can be seen in figure 3.2.

The Bickmore agent (Laura) tries to improve the attitude of the user towards exercising. This agent is more focussed on dialogues. The responses of this agent are fixed; all possible conversations are stored as a dialogue tree. Users select their answer by clicking on the button with their answer, after which the system checks the dialogue tree and continues with the dialogue from there. In a experiment of Schulman and Bickmore [83], participants were asked to speak their choice, but to restrict their utterances to the choices given. This however, was a Wizard-of-Oz arrangement (unknown to participants, a researcher listened via a microphone from an adjacent room, and selected the response that matched their utterance.). The agent could deliver output as synthesized speech with synchronized nonverbal behaviour [83, 12]. An image of the Bickmore agent can be found in figure 3.3.

Figure 3.1: A user interacting with Rea

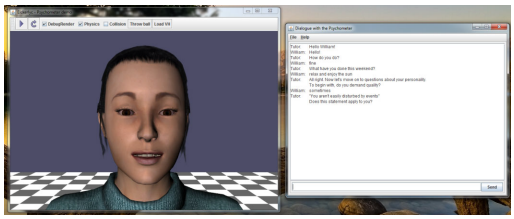


Figure 3.2: Psychometer

ECA in the SmarcoS system One of the versions of the SmarcoS system that will be used in this thesis includes an ECA. It is well known that the use of an ECA has a positive effect on user experience [11], for example in persuasive systems. This is a good reason to include an ECA in one of the versions of the SmarcoS system. The system uses a smartphone as the main way of communication. Therefore, the used ECA should run on a smartphone. Using a full 3D virtual human would be too heavy to use on a smartphone in terms of processing power and battery usage, and it would be unclear on a relatively small screen of a mobile phone. A light-weight animation embodiment is used for this; PictureEngine. This enables us to use the Elckerlyc platform on a mobile phone. We will now discuss Elckerlyc, after which we will tell something about the PictureEnging.

Figure 3.3: Bickmore agent

The Elckerlyc platform is a Behaviour Markup Language (BML) realizer. It can generate behaviours of virtual humans real-time. “BML provides abstract behaviour elements to steer the behaviour of a virtual human” [82]. How these abstract behaviours will be displayed on the embodiment, a BML realizer is free to choose.

The PictureEngine is a lightweight graphical embodiment that uses a collection of

2D images in order to display the ECA [51]. It uses layers to display different parts of the ECA; therefore they can be in different states. By using this layer approach, all parts of the ECA can be manipulated independently, which combined generates different expressions. This also has some limitations; any movement of the entire ECA is a problem. However, due to the screen size of mobile phones locomotion is impractical. But smaller movements such as nodding, shaking and tilting of the head are also problematic. PictureEngine allows the use of animation, because there are cases where an ECA had to display some motion in order to be believable. These animations are defined using a simple XML format. This format also allows a synchronization point to be included in the specification between two frames of an animation [51]. The PictureEngine also uses a binding, which allows a combination of a BML behaviour class and possibly some constraints to be mapped to a certain PictureUnit. Finally, PictureEngine also provides a rudimentary lipsync [51].

The system can use the internal text-to-speech (TTS) system that Android provides. However, this has the problem that no timing information for utterances can be obtained with the Android TTS system. Therefore, the BML scheduler can not use synchronization points within utterances. This is the main reason the PictureEngine on Android does not support lip-synchronisation. The PictureEngine also uses subtitles of the spoken text, because the high chance that users might have trouble hearing the spoken text.

We explained shortly what an ECA is, and gave some examples of existing ECA's. After this we discussed the mobile phone version of the ECA the SmarcoS system uses. We explained how this is facilitated and discussed some limitations of the ECA.

In this chapter we looked at behaviour change support systems in general, important psychological theories for behaviour change support systems, the various generations of behaviour change support systems we can distinguish, and elements of a behaviour change support system. Secondly, we discussed Embodied Conversational Agents. This theory will serve as background information about the system. Secondly, when discussing the results of the evaluations of the SmarcoS system, we will use this theory as a reference for explaining some of the results. In the next section we will look at evaluating systems.

4 Evaluating systems

Whenever you make a system it is important to evaluate it. Does your system work like you thought it would? What improvements can you make? Is your system accepted by its users? How is its usability? Is it easy to use, easy to learn? How does the user experience you system? Which questions you want answered of course influences your evaluation, the system you have developed and its components also influence it. In this chapter we will give an overview of the questionnaires that we looked at for evaluating the SmarcoS system. We will shortly say something about each of the questionnaires, what is its origin?, what does it measure?, how does it measure it? and how reliable is it? This overview further on will be used as basis for the evaluation of the SmarcoS-office worker system. We will use whole questionnaires or parts of them in the evaluation, how this evaluation will be composed can be found in section 6.2.3.

4.1 Usability

Usability is, among other aspects, a measurement for how easy it is to use a product and how easy it is to learn how to use it. The ISO definition is as follows: "The extent to which a product (service or environment) can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [45, 76]. To measure the usability of a product there are many questionnaires. One of the questionnaires that is often used to measure usability is the *System Usability Scale (SUS)*.

SUS is a simple and reliable, ten-item scale that will give a global view of subjective assessments of usability of a system. SUS was created from a pool of 50 potential questionnaire items. Items leading to the most extreme responses from the original pool of 50 items were selected. Questions from this questionnaire cover a variety of aspects of system usability, such as the need for training, support and complexity. Items are answered on a 5-point Likert scale. Respondents should be asked to record their immediate response to each item, rather than thinking about items for a long time. Based on the ratings given by participants a single score (ranging between 0 and 100) is calculated which indicates the usability of a product [14]. The System Usability Scales can be found in appendix A.1.

SUS has several attributes that make it a desired scale for a broad range of people. First of all, the survey is technology independent. This makes it flexible to assess a wide range of technologies. Secondly, SUS is quick and easy to use by participants as well as administrators. Third, the single score that is provided by the survey is easily understood. Finally, the survey is not protected by trademark, patent or copyright, which makes it a cost effective tool [6]. Reliability of SUS was calculated at different times, ranging from Cronbach's alpha = 0.85 till 0.911. Factor analysis results show that the SUS questionnaire reflects participants' estimates of the overall usability of an interface, regardless of the type of interface[6].

What makes a good SUS score? It is theorized that a product is at least passable with a SUS score above 70. Better products score in the high 70s to upper 80s, while superior products score better than 90 (figure 4.1) [6].

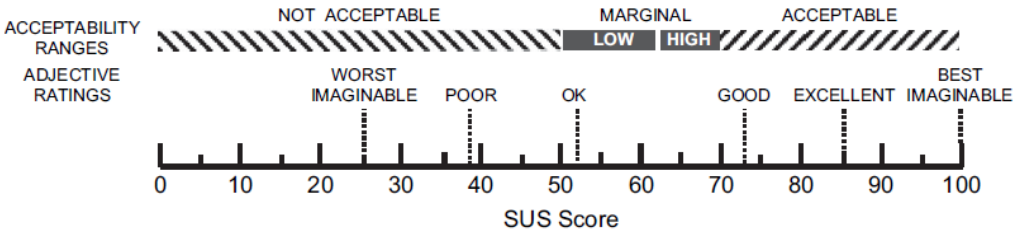


Figure 4.1: Adjective ratings, and acceptability of the overall SUS score

4.2 Questionnaire for User Interface Satisfaction

The *Questionnaire for User Interface Satisfaction (QUIS)* measures the user's subjective rating of a human-computer interface. The original questionnaire consisted of a total of 90 questions. Different versions have been made during the years to improve the original version. It had to be shortened to improve the percentage of completed questionnaires, each successive version having fewer items while maintaining a high reliability. Various versions have been made. The reliability of the discussed version (version 5) is high, Cronbach's alpha = 0.94. This version has 27 questions measuring five constructs; overall reaction to the software, screen, terminology and system information, learning and system capabilities [19]. The questionnaire is measured on different scales (the first six questions are 10-point semantic differentials, remaining questions are answered on a 10-point scale (for the questionnaire see appendix A.2). This study [19] established external validity, the QUIS has good discriminability in the overall reaction ratings for like and dislike. However, no attempt to establish any construct or predictive validity was done.

4.3 User experience

User experience (UX) is associated with a variety of meanings. They range from traditional usability to beauty, hedonic, affective or experiential aspects of technology use. Hassenzahl [36, p.12] defines it as: "a momentary, primarily evaluative feeling (good-bad) while interacting with a product or service". There are many models to describe the nature of user experience; they all focus on well-being as an outcome of human-product-interaction, and not on performance of a product. This asks for an enrichment of traditional quality models with concepts such as fun, hedonic value or playfulness. It calls for a holistic perspective, and shifts attention from the product and materials to humans and feelings. User experience encompasses all aspects of interaction with a product and is subjective. Therefore, the actual experience of a product can differ from intended experience by the designer. UX is subjective since it is a consequence of a user's internal state, the characteristics of the designed system and the context within which the interactions occur [34, 36, 39].

To measure UX two dimensions play an important role: pragmatic quality (PQ) and hedonic quality (HQ). Pragmatic quality is connected to the perceived ability of the product to achieve "do-goals". These are goals such as "finding a book in an online-bookstore", "making a telephone call", or "setting-up a webpage". These are all behavioural goals. Attributes that can be linked to pragmatic quality are "useful", "supporting", "clear", and "controllable". Pragmatic quality focusses on the product, its utility and usability in relation to potential tasks [34, 35, 36].

Hedonic quality is connected to the perceived ability of the product to achieve "be-

goals". These are goals such as "being related to others", 'being competent", or "being special". These are all goal related to the users' self. It has been shown that hedonic qualities play a role in UX. Attributes that can be linked to hedonic quality are "exciting", "impressive", "outstanding", and "interesting". Hedonic quality focuses on the self. Why does someone use this product, and not the other? Also, more general human needs come into play, such as a need for personal growth, for novelty and change, and self-expression. The hedonic function can be further subdivided into stimulation (HQ-S) and identification (HQ-I). Stimulation is the part of hedonic qualities that focuses on personal development, qualities that provide new impressions, opportunities and insights. Identification is the part of hedonic qualities that focuses on expressing one self, qualities that help with self-expression [34, 35, 36, 37].

A well-known technique for evaluating objects, and measuring how people perceive them, is the semantic differential scale. It has various advantages: the usability engineer does not require a special training for using the differential, the participants can quickly and easily fill it in, and the statistical analysis is straightforward [37]. *AtrakkDiff2* is such a scale for measuring UX, consisting of 28 word-pairs answered on a 7-point scale (for the questionnaire see appendix A.3. It focuses on attractiveness of interactive products and evaluates the following dimensions: pragmatic quality (PQ), hedonic quality - stimulation (HQ-S), hedonic quality - identity (HQ-I), and attractiveness (ATT). It states that hedonic and pragmatic qualities are independent of one another and contribute equally to the rating of attractiveness. It can give insight in how people experience a product, and what qualities should be improved to enhance this experience. Reliability is shown for hedonic quality - stimulation (Cronbach's alpha 0.79 - 0.90), hedonic quality - identity (Cronbach's alpha 0.73 - 0.83) and pragmatic quality (Cronbach's alpha 0.83 - 0.85) [38], and shown in other studies as well [35].

4.4 Technology acceptance

The acceptance of a system or product by its users is important. Without accepting the system a user will not use it, no matter how good the product is, or how useful. In this section we will first discus the *Technology Acceptance Model (TAM)* with two of its most important constructs. Finally we will discuss the *Unified Theory of Acceptance and Use of Technology (UTAUT).*

4.4.1 Technology Acceptance Model

The *Technology Acceptance Model* is used to address why users accept or reject information technology and how user acceptance is influenced by system characteristics. It can explain why a system is unacceptable to some users, but also improve understanding how we can gain user acceptance through system design [24].

The technology acceptance model is based on the *Theory of Reasoned Action (TRA)* of Fishbein and Ajzen's [29]. TAM uses TRA as a theoretical basis for specifying the causal relationship between: perceived usefulness and perceived ease of use, and users' attitudes, intentions and actual usage (see figure 4.2) [25, 24]. TAM theorizes that the behavioural intention of a person to use a system is determined by two beliefs: perceived usefulness and perceived ease of use. It also says that the effects of external variables are mediated by those two constructs. Perceived usefulness is influenced by perceived ease of use because, if all other things are equal, the easier system is more useful. Therefore, perceived usefulness and perceived ease of use are hypothesized to be fundamental determinants for user acceptance. Perceived usefulness can be defined as "the degree to which a person believes that using a particular system would enhance his

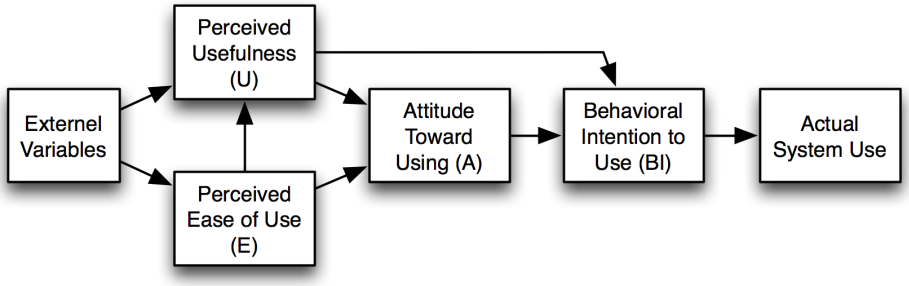


Figure 4.2: Technology Acceptance Model

or her (job) performance". And we can define perceived ease of use as "the degree to which a person believes that using a particular system would be free of effort" [23].

People tend to use an application to the extent they believe they will benefit from it, whether it is in their daily lives or in their job. Even if a user believes that a system is useful, but he also believes that the system is too hard to use, then the benefits of using the system are outweighed by the effort of using the system and the user will not use it. Davis [23] used a step-by-step process to develop scales measuring perceived usefulness and ease of use with high reliability and validity. For each construct 14 candidate items were generated based on definitions from literature. From these items the 10 items that fit the definitions of the constructs best were selected for each scale. This version was tested, and reliability and validity were calculated to be both high. Since it is important to keep scales as brief as possible in a testing situation, these 10-item scales were adapted to six-item scales. The items are answered on a 7-point Likert scale (for these scales we refer to appendix A.4.1). Items that contributed least to the reliability were omitted. The resulting scales were again tested and reliability was measured. Cronbach's alpha was 0.98 for perceived usefulness and 0.94 for perceived ease of use. Both scales exhibited high convergent, discriminant, and factorial validity [23]. Hendrickson [43] did a test-retest of the reliability, which confirmed the findings of [23].

4.4.2 Technology Acceptance Model 2

In 2000, TAM was extended to TAM2 (see appendix A.4.2 for the questionnaire) to include additional key determinants of perceived usefulness and usage intention constructs, and to understand how the effects of these determinants change with increasing user experience over time with the system (see figure 4.3) [90].

Original measurements with TAM2 showed high reliabilities with Cronbach alpha coefficients exceeding 0.80. Construct validity was strongly supported both by principal components analysis and by an analysis of the multitrait-multimethod matrix. TAM2 provides a detailed account of the key forces underlying judgments of perceived usefulness, explaining up to 60% of the variance in this important driver of usage intentions.

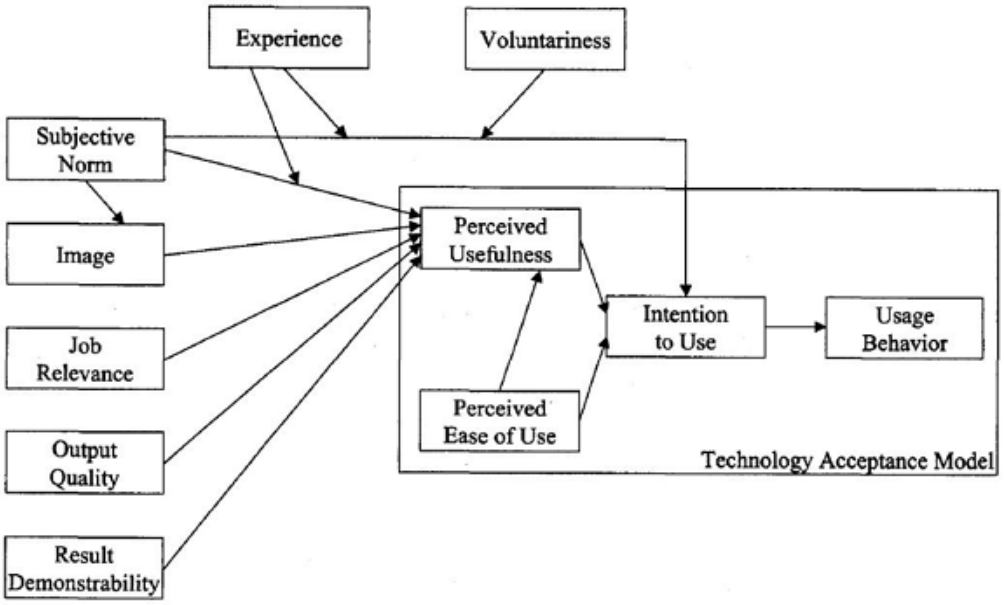


Figure 4.3: Technology Acceptance Model 2

4.4.3 Unified Theory of Acceptance and Use of Technology

There are many competing models in the world of information technology acceptance, and the previously mentioned TRA, TAM and TAM2 are just a few of those out there. They routinely explain over 40% of the variance in intention to use the technology. One stream of research has focused on individual acceptance of technology by using intention or usage as a dependent variable; other streams have focused on success at the organizational level and task-technology fit among others. All streams make important contributions to the literature. The *Unified Theory of Acceptance and Use of Technology (UTAUT)*, combines these theories based upon conceptual and empirical similarities across model. It is formulated with four core determinants of intention and usage, and up to four moderators of key relationships. The questionnaire is answered on a 7-point Likert scale, for the whole questionnaire we refer to appendix A.4.3.

As can be seen in figure 4.4 these four core determinants of intention and usage are: Performance Expectancy, Effort Expectancy, Social Influence and Facilitating Conditions. In this model performance expectancy is defined as "the degree to which an individual believes that using the systems will help him or her to attain gains in job performance" [91]. The effort expectancy is defined as "the degree of ease associated with the use of the system" [91]. Social influence is defined as " the degree to which an individual perceives that important others believe he or she should use the new system" [91]. Finally, facilitating conditions are defined in the model as "the degree to which an individual believes that an organizational and technical infrastructure exists to support use of the system" [91]. The four moderators of key relationships are; gender, age, experience and voluntariness of use. UTAUT is tested and cross-validated, these test provide strong empirical support for the model. UTAUT was able to account for 70% of the variance in usage intention, which is better than any of the original models used to compose UTAUT [91]. It is also shown that the UTAUT tool is able to withstand translation and to be used cross-culturally, outside its original country and language of origin [72].

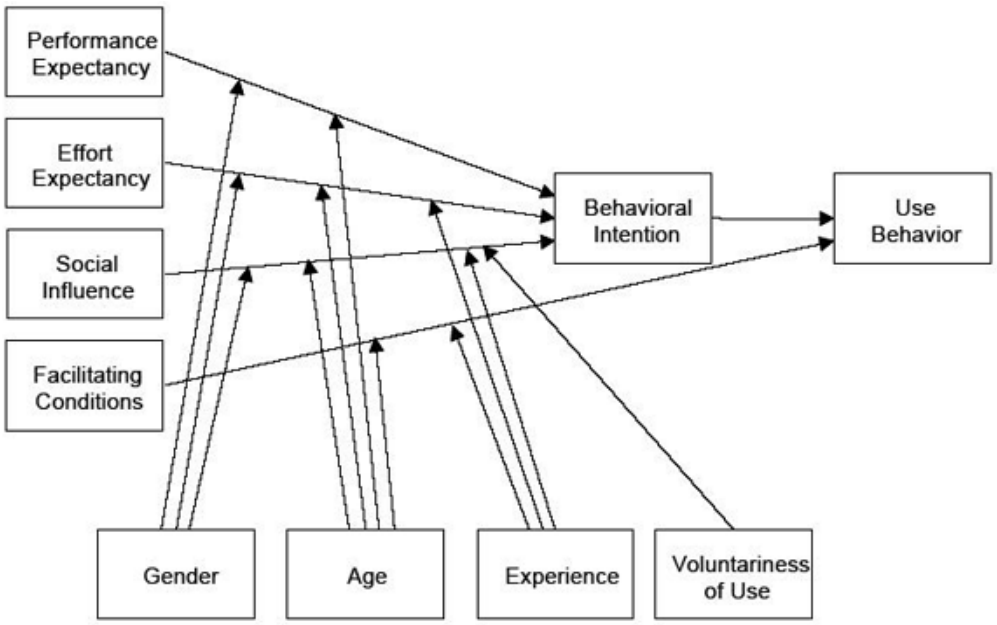


Figure 4.4: Unified Theory of Acceptance and Use of Technology

4.5 Robot acceptance

There are many questionnaires about ECA's and their personality. However, there is not a good questionnaire concerning the acceptance of ECA's by users. Robots come most close to ECA's when we look at the acceptance of them, especially since they have improving social abilities, and more functionality. Therefore, we will look at robot acceptance.

4.5.1 Heerink

Both improving social abilities and increasing functionality of robots influence the acceptance of robot interfaces. There are several research groups focusing of robot acceptance, among which a Dutch group. This group around Marcel Heerink focuses on two main concepts being: social abilities for robots, and user acceptance of robots. They developed a questionnaire combining these concepts creating a questionnaire that focuses on the acceptance of robots with social abilities.

The UTAUT model is a sound basis to start, due to its extensive validation and the potential applicability of the model to human-robot interaction as is indicated by De Ruyter et al. [26]. The Dutch research group made adaptations to the UTAUT because of three reasons. First, participants had difficulty indicating the level to which they agreed with statements. Therefore, statements were adapted to questions. Secondly, these questions were asked by an interviewer instead of read by the participants since some of the participants had trouble reading. Finally, UTAUT was adapted to fit the test setting better, since UTAUT is originally developed for using technology at work. Other questions were added concerning trust, social abilities, computer experience and one question concerning the extent to which people felt comfortable. Questions were answered on a 5-point scale (see appendix A.5.1). Cronbach's alpha was calculated for

all UTAUT constructs to see if they were consistent. All constructs had a Cronbach's alpha of 0.86 or higher except for social influence and anxiety [41].

4.5.2 Almere model

In another study this research group adjusted their questionnaire because of the low explanatory power, and because it insufficiently indicated that social abilities contribute to the acceptance of a social robot. They carried out several studies focusing a possible constructs to add. Perceived Enjoyment, Perceived Sociability, Social Presence and Perceived Adaptability were found and added. Anxiety and Attitude toward using the technology were also added, although they are not part of the UTAUT model. The resulting questionnaire (Almere model) now measures in 41 questions the constructs of: Anxiety, Attitude, Facilitating conditions, Intention to use, Perceived adaptability, Perceived enjoyment, Perceived ease of use, Perceived sociability, Perceived usefulness, Social influence, Social presence, Trust and Use/Usage. For the whole questionnaire we refer to appendix A.5.2. Reliability was tested and all constructs were shown to be reliable (Cronbach's alpha of 0.7 or higher) [42]. The questionnaire was further tested in other experiments, and constructs are shown to be reliable in these studies as well [40].

4.5.3 GODSPEED

Finally, we would like to discuss the GODSPEED questionnaires from Christoph Bartneck [9]. The questionnaires can be found in appendix A.5.3. These short questionnaires are an attempt to standardize measurement tools for human robot interaction. To form these questionnaires they reviewed relevant literature on the five key concepts of anthropomorphism, animacy, likability, perceived intelligence, and perceived safety of robots. The most important criteria of service robots lie within the satisfaction of their users, unlike the criteria for industrial robots in which it is far more important how many pieces they can process and what their accordance is with quality standards. Because user satisfaction is more important we need to measure the perception of the user of service robots.

GODSPEED measures anthropomorphism, animacy, likability, perceived intelligence, and perceived safety of robots. We will shortly explain all concepts.

Anthropomorphism refers to the attribution of human form, human characteristics, or human behaviour to nonhuman things. Or shortly said, how humanlike we think a non-human being is. Bartneck [9] found that the questionnaire of Powers and Kiesler [78] was best suited to measure anthropomorphism. They adapted the six items of this questionnaire into 5-point semantic differentials (all questionnaire of GODSPEED are semantic differentials, to improve coherence): fake–natural, machinelike–humanlike, unconscious–conscious, artificial–lifelike, and moving rigidly–moving elegantly. Studies using this scale report internal consistency reliability of Cronbach's alpha of 0.856 or higher [9].

Animacy can be seen as how lifelike the robot is. Bartneck [9] found that the questionnaire of Lee et al. [57] best represented this construct. Again the questionnaire was transformed into semantic differentials: dead–alive, stagnant–lively, mechanical–organic, artificial–lifelike, inert–interactive, and apathetic–responsive. One study [8] used this questionnaire and reported a Cronbach's alpha of 0.702.

Likeability, or the positive first impression of a person often leads to more positive evaluations of that person. Bartneck used five items from Monahan [65]: dislike–like, unfriendly–friendly, unkind–kind, unpleasant–pleasant, and awful–nice. Bartneck reports two studies using this questionnaire all having internal consistency reliability of Cronbach's alpha of 0.842 or higher [9].

Perceived Intelligence is quite straightforward. Warner and Sugarman's [93] scale for intellectual evaluation was used for this construct. It consists of five items: incompetent–competent, ignorant–knowledgeable, irresponsible–responsible, unintelligent–intelligent, and foolish–sensible. Multiple studies [8, 7, 10, 50, 73]used this questionnaire, all reporting Cronbach's alpha's of 0.75 or higher.

Perceived safety describes the user's perception of the level of danger when interacting with a robot, and therefore also the level of comfort of the user during the interaction. Bartneck could not find a suitable questionnaire for rating the safety of robots. The items of this construct are based on [54, 55]. The items are: anxious–relaxed, agitated–calm, and quiescent–surprised. No reliability is reported for this specific scale. It should be noted that there is a certain overlap between anthropomorphism and animacy.

4.6 Source Credibility

Source credibility is the attitude toward a source of communication held at a given time by a receiver. In general, research supports the proposition that source credibility is a very important element in communication processes, whether the goal of the communication effort is persuasion or understanding. People are more likely to be persuaded when the source is perceived as credible and is presented that way [64, 85].

Although source credibility it mostly seen as a human-human interaction, we will use this measurement for the SmarcoS system. The SmarcoS system is meant to represent a coach motivating the user to be more active. Therefore, we could argue that the system represents a person and source credibility can be applied although this person is only represented by text, or an ECA (Embodied Conversational Agent 3.2) [47].

Originally source credibility was seen as a one-dimensional attitude the receiver had about a source. This changed when two lines of research began promoting it as a multidimensional attitude. The multidimensionality of the construct itself was already noted in classical times as well. An example of this is Aristotle, who suggested that ethos (or source credibility) had three dimensions: intelligence, character and good will. Source credibility is a subset of a much larger construct of person perception [64].

A 5-point Likert-type scale was made by McCroskey [63] in which two source credibility dimensions were measured, authoritativeness and character. Originally authoritativeness consisted of 22 items and character of 20 items. McCroskey conducted several studies to develop and test this credibility instrument, and created two constructs with six 7-point semantic differential scales (see appendix A.6.1). Several years later this 7-point semantic differential instrument was revised and extended to an instrument containing five dimensions each consisting of three bipolar constructs. These include Sociability, Character, Competence, Composure and Extroversion and can be found in appendix A.6.2. Both scales show high internal reliability (the two times six-item scale had Cronbach's alpha values of 0.93 (authoritativeness) and 0.92 (character), while for the 15 item scale different alpha's are reported between 0.68 and 0.96). Construct validity was shown for the two times six-item scale, while construct validity for the 15 item scale remains questionable, because it has not always factored into five dimensions. Their use by researchers during the years indicates their predictive and construct validity. The twelve item version is used more [63].

4.7 Coaching behaviour and quality

A way to measure coaching behaviour and its quality is by using the Coaching Behaviour Scale for Sport (CBS-S). It is based on qualitative research with coaches and athletes providing the theoretical base. The objective behind the development of the CBS-S was to provide a measurement instrument that closely represented coaching behaviours in various sports, at various levels. The scale is easier to use than for example the Coaching Behaviour Assessment System (CBAS), or an adapted version of that instrument, since it is a 7-point scale and the CBAS is an observation instrument [22].

Developing the scale, 75 item for the CBS-S were derived from a series of qualitative studies with coaches and athletes, and input of the Institute National du Sport et de l' Education Physique. All items were drafted into questionnaire format and reviewed for readability and face validity by eight academics and three coaches. This questionnaire was then completed by 105 rowers [22]. Afterwards, all items underwent an exploratory factor analysis. This resulted in 37 items forming six factors. These factors were; Technical Skills (8 items about coaching feedback, demonstrations and cues), Goal setting (6 items assessing the coach's involvement in the identification, development, and monitoring of goals), Mental preparation(5 items assessing the coach's involvement in helping the athlete be tough, stay focused and be confident), Personal Rapport (7 items assessing the approachability, availability, and understanding of the coach), Physical Training (8 items about the coaches' provision of physical training and planning for training and competition) and Planning and Negative Personal Rapport (3 items describing the coach's use of fear, yelling when angry, and disregarding the athlete's opinions). These 37 items were used for further developing the CBS-S [22].

A second study was done with a more diverse and larger sample of athletes (N=205). Athletes were asked to complete the questionnaire now containing only the 37 items acquired in the first study (for the questionnaire we refer to appendix A.7.1). The items were again submitted to the factor analysis, tested on reliability (internal consistency and test-retest reliability) and validity (factor validity) [22]. The same six factors emerged from the analysis. Each factor had high item loadings, indicating strong factor validity. All constructs demonstrate very high internal consistency with Cronbach's alpha coefficients of 0.85 or higher. Test-retest reliability was based on a small and convenience sample (N=67). All positive constructs demonstrated adequate test-retest reliability, the negative construct (Negative Personal Rapport) was lowest at $r = 0.49$ [22].

Philips has adapted this scale for measurements of the DirectLife coaching system (see appendix A.7.2); however, there is no documentation about it.

In this chapter we gave an overview of several questionnaires that could be used for evaluation of the SmarcoS-system. We stated what the origin of the questionnaire is, what it measures, how it measures it and how reliable it is. We will use several of these questionnaire or parts of them in the evaluation of the SmarcoS-office worker system. For more information on which questionnaires are used and the reasoning behind this, we refer to section 6.2.3.

5 Evaluation SmarcoS-diabetic

In this chapter we will discuss the evaluation of the SmarcoS-diabetic system. This evaluation is conducted in collaboration with Evalan. Evalan is an "innovation company with focus on telemetry solutions and M2M (machine to machine) services". They provide full-service telemetry solutions to industrial companies, research facilities, healthcare institutes and private consumers. To support the delivery of these telemetry services, they develop mobile devices and sensor units, data management systems, data processing algorithms and user interfaces on various platforms. Evalan works often in cooperation with international technology partners and universities [2].

We will first explain the SmarcoS system and the version of the system tested in this chapter (SmarcoS-diabetic (SD)). This version is based on the second use case presented in the introduction in which the target group are diabetics. After this we will discuss the methodology. We will end by presenting the results and discussing them.

5.1 The SmarcoS system

This thesis looks at three versions of the SmarcoS system. The versions focus on different target groups, as already discussed in the introduction (section 2). Because they focus on different target groups, the versions of the system share some features, but they also differ at some points. In the following section we will first discuss the features that all versions of the SmarcoS system share. Secondly, we will look at the version focused on diabetics, which we call SmarcoS-diabetic (SD). The other two versions focused on office workers will be discussed in the next chapter in section 6.1.

5.1.1 Shared Basis

All versions of the system used in this thesis have the goal to help people achieve a healthy and balanced lifestyle, therefore they support behaviour change. The system accommodates this behaviour change by giving feedback on occurring situations. A common goal for all versions of the system is to motivate users to live a more active life. This is done in all systems by giving feedback on activity data that is measured.

As basis for the system Philips DirectLife is used [1]. The Philips DirectLife system stimulates users to improve their activity level, and be physically more active. To achieve this, DirectLife measures the activity level of the user and gives feedback on this by sending an e-mail. The activity level of participants is measured using a triaxial accelerometer (see figure 5.1a), named activity monitor (AM). The SmarcoS system uses the AM as a way to measure the activity of participants. Furthermore, the SmarcoS system uses the Philips DirectLife system to process the activity data.

The activity monitor can be seen in figure 5.1a, actual size. Participants have to carry this monitor with them throughout the whole day. It can be worn at different positions shown in figure 5.1b. When cycling, the activity monitor should be placed in the sock of the user. It is also possible to swim with the activity monitor, since it is waterproof to up to three meters under water. To extract the activity data, the AM has to be connected to a computer using a magnetic USB connector. The AM is battery charged and will automatically charge when connected to the computer. As soon as the AM is connected to the computer (docked), it will synchronize the activity data with the data on the server. Secondly, it will open the DirectLife website, where the synchronized activity data can be seen. The activity monitor itself has nine small green

(a) Triaxial accelerometer of DirectLife

(b) Wearing positions of the activity monitor

Figure 5.1: Triaxial accelerometer of DirectLife, with corresponding wearing positions

lights, which provide an indication of the amount of activity done that day (depending on the number of lights that burn). For all participants a personal activity goal is set. This goal increases over time to stimulate more physical activity. The way the activity goal is set differs per version and will be discussed in section 5.1.2 and section 6.1.

The SmarcoS system is a rule based system, this means that a set of predefined rules is used to decide when to send a message, and which message is sent to the devices. We will discuss the rules for each version in section 5.1.2 and section 6.1. All versions of the system are able to give feedback using an application on a smartphone. When a new message is received a notification is given by the smartphone. Messages and activity level (including history of the activity level in percentages) can be viewed on the smartphone. Feedback on the amount of activity already done that day is given by all systems.

5.1.2 SmarcoS-diabetic (SD)

The SmarcoS-diabetic version differs from the SmarcoS-office worker version at four main points. First of all, this version also monitors medication intake. Secondly, it has a separate computer application. Thirdly, it has a different set of rules, and finally, the way the activity goal is set differs. We will now discuss all the above points and the smartphone application of this version

Pill dispenser An important feature in the SmarcoS-diabetic version of the system is the pill dispenser (see figure 5.2). The pill dispenser can register being opened by someone to take a pill. It monitors real-time if medication is taken [2]. Each time a participant takes his medication, this "medication event" is registered, sent and stored in a central database within half a minute. In this version of the system the event is also sent to the DirectLife database.

Figure 5.2: Pill dispenser (sensemedic)

The times on which the participants normally take their medication are set beforehand. Participants then

have a 30 minute timeframe meaning 30 minutes before this time and 30 minutes after this time in which they can take their medication. After this time frame a message with a reminder is sent to the participants to take their medication.

Smartphone application As can be seen in figure 5.3 the smartphone application has three icons, each representing a function. We will discuss each of these icons, combined with their functionality.

Figure 5.3: Smartphone Application

First of all: the top icon. When selecting the top icon, the user can see his activity overview of the last six days (this is shown as the total percentage of the goal for that day). The icon itself shows the last known activity data progress in percentage (in this case 19%) towards the activity goal of that day. If the user has not yet docked his activity monitor, it states "please dock".

The second icon concerns medication intake. Depending on the number of moments in a day that users have to take medication, a different number of small clocks are shown, each representing a medication intake moment. The clocks become red if users forgot to take their medication, orange when they were too late or too early and green if they did take their medication on time. The clocks are grey when the medication moment has not yet passed. When selecting the icon the user can see his medication history of the last six days (also shown in the form of coloured clocks). As can be seen in the example of figure 5.3, this person has three moments on a day to take his medication, he missed the first two.

The last icon (bottom of figure 5.3) shows how many unread messages the user has (in this case there are no unread messages). By selecting this icon the user is shown his inbox of messages. When selecting one of the messages the user can read the whole contents of the message. Like in an e-mail inbox, a subject of the message is shown as well as the time the message is sent. Opening a message in the SD version just shows the content of the message in text.

Computer application The SD version of the system also has a computer version of the smartphone application. This way, data is presented on two different devices but basically in the same way, forming one system. It has the same functionality as the smartphone application. An image of the interface can be found in figure 5.4. In this interface we see the same icons that are used in the smartphone application and one extra icon (most left icon) to go to the overview. Underneath the icons we see the overviews, from left to right, activity history of the last week, part of the message inbox and medication history. We will shortly give some information about each item.

There are two icons which can be clicked on (most left icon, and the second of right icon), when hovering over those icons with the mouse it highlights them to signal this.

The most left icon reads 'Go to overview'. When you have left the overview that is visible in figure 5.4 you can use this button to go back to the overview.

Figure 5.4: Computer Application

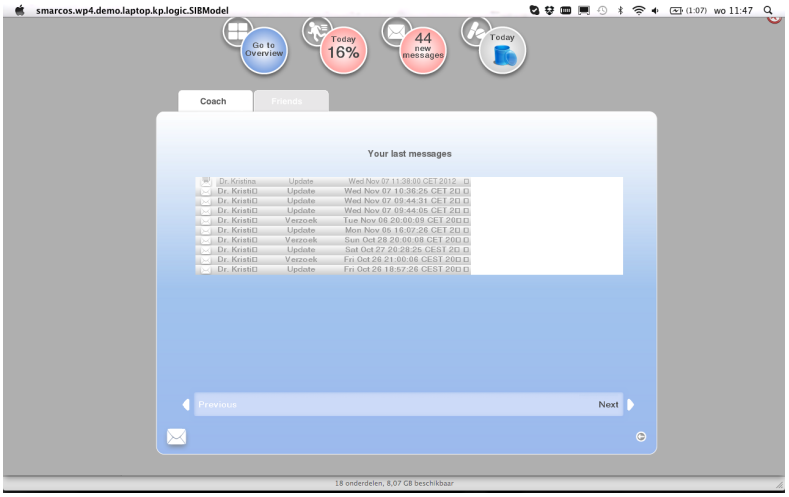


Figure 5.5: The message inbox in the computer application

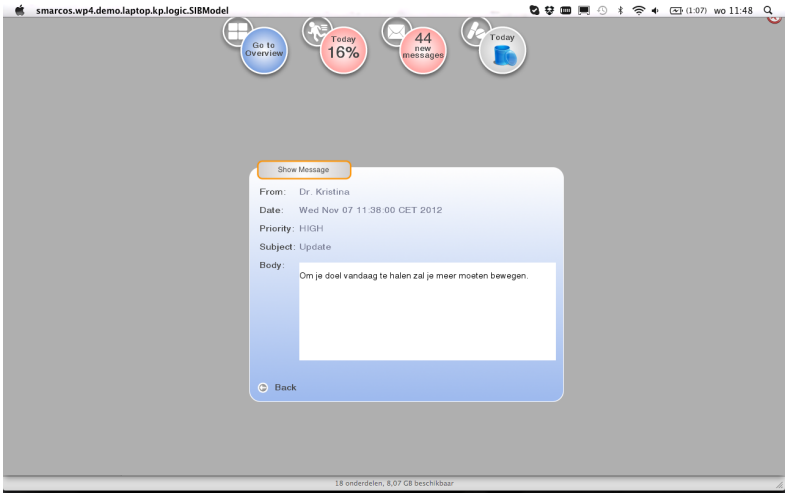


Figure 5.6: A message in the computer application

The second left icon shows the last known activity data progress in percentage (in this case 72%) towards the activity goal of that day. Or, if the user has not yet docked his activity monitor, it states please dock. As already said, the activity history of the last week can be seen in the most left overview.

The third icon shows how many unread messages the user has (in this case there are two new messages). The overview in the middle of figure 5.4 shows part of the messages inbox. Unread messages are shown in bold. It is possible to go to the whole message inbox (figure 5.5) by clicking this icon, or by clicking on the message block below it. The message itself (figure 5.6) can then be read by clicking on it.

The most right icon concerns medication intake. Each clock represents a medication intake moment as in the smartphone application. As can be seen in the example of figure 5.4, this person has five moments on a day to take his medication, he was on time the first three moments, the fourth he took too early or too late, and the last still has to come. The most right overview shows the medication history of this week.

Rules The rules of the SD version provide feedback based on all input data (in this case, the activity data and the data of the pill dispenser). There are four rules in this version of the system for providing feedback. These rules we will now shortly discuss, for each rule we will give an example sentence that is send to the user. The texts that are sent to the smartphone application in this version are not the same as in the other two versions of the systems (see section 6.1).

- **NoMedicationIntake:** When the medication event is missed and no medication reminder is read, send a "medication reminder".
  "Did you already take your medicine?"

- **Medication_100:** When medication is taken three days in a row correctly, and no medication goal reached messages is send, send a "congratulation message".
  "Congratulations, you have taken your medication on time in the past 3 days. Good job."

- **ActivityGoalReached:** When the last activity achievement is more than 100% and no "activity goal reached" messages is send, send a 100% "activity message".
  "Congratulations, you have reached your activity goal for today!"

- **ActivityLowOffice:** When the last activity achievement is less than 30% and less than two "activity low" messages are send, send an "activity low message".
  "Compared to your daily routine, your activity level is low. Consider doing something physically active in the following hours."

Setting the activity goal During the evaluation of this version, activity goals were set by researchers before starting the evaluation. This has the advantage that participants do not need to do an assessment week, in which their normal activity level is determined. The main disadvantage is that the set goal might not be appropriate. This can lead to participants receiving messages that might frustrate them, because the goal level of activity is too high to reach and they keep getting reminders to be more active. Or the goal might be too low, which leads to receiving messages that participants have reached their activity goal too easily. The latter is less of a problem except that the system is not fulfilling its goal; motivating the user to be more active.

Generation of behaviour change support system This version of the SmarcoS system belongs to the third generation of behaviour change support systems, since it runs on a smartphone, and gives feedback that is tailored on the situation.

5.2 Methodology

In this section we will first provide information about the participants, after which we will discuss the procedure. Finally, we will tell something about the materials and shortly explain the data-analysis.

5.2.1 Participants

The participants in this study were five type II diabetics diagnosed 1.5 to 2 years ago. Participants were between 30 and 52 years old (M=41.6 year, SD=7.9). We had two male participants and three female participants. Four participants own a smartphone and use it daily; the other participant did not have any experience with a smartphone. Participants indicated that they used a computer for different activities namely work (4/5), hobby (3/5), internet (2/5), administration and data storage (1/5).

5.2.2 Procedure

At the beginning of the evaluation, each participant was visited at home. Each participant received the devices (activity monitor, pill dispenser, applications for smartphone and computer). If necessary a smartphone was provided. Note; not all participants received the activity monitor during their home visit, since activity monitors did not arrive yet. If necessary, they were sent the next day by mail.

During this visit the system and applications were explained to the participants. Consent forms were given combined with an information letter that included the objectives of the test and practical organizations. The informed consent included an agreement on returning the devices in good conditions. Consent forms were signed by the participants after going through them with the researcher. An introduction questionnaire was given to fill out. These questions were about age, gender, previous participation of the participants in an earlier stage of the project, how long ago the diagnose diabetic was made, their computer experience and their smartphone experience. Applications were installed and an explanation of the functionality was given. After verifying that participants understood the applications, diaries (see section 5.2.3) were given with a short explanation. Finally, we asked the participants about their expectations. Each introductory visit took about 1.5 hour.

Participants used the SD system for a whole week (7 days). They were asked to regularly dock their activity monitor, so the system could give feedback on this data. The SD system provided feedback about their activity level and medication intake.

Each day participants filled out a day of the diary (see materials 5.2.3, diary and appendix B).

Shortly after the test week, an interview took place with each participant. For a description of this interview we refer to the following section.

5.2.3 Materials

In this section we will provide information about the materials used in this evaluation. We will discuss the diary the participants were asked to fill out, the log-data that was gathered, and the interview that was held afterwards.

Diary During this study participants were asked to fill in a diary. This diary consisted of a short questionnaire for every day, each having its own subject. The aim of these questionnaires was to gather information about the users' interactions with the system that cannot be gathered from the database or logs. The diary also contained phone numbers and e-mail addresses of the researchers, so participants could contact them in

case of questions, remarks or problems with the system. The following subjects were covered:

- **Day 1:** How active were you today? These questions were about activity messages.
- **Day 2:** What do you do to remember to take your medication? These questions were about medication messages.
- **Day 3:** What sources do you use to gain information about healthy living?
- **Day 4:** When do you take you medication? These questions were also about medication messages.
- **Day 5:** Different functions of the system. These questions were about the functionalities of the system, what did the user use and on which device (smartphone/computer).
- **Day 6:** Words that are applicable to the system. Participants were asked to assign words from a list to different parts of the system. Also the word pairs from Attrakdiff questionnaire (see section 4.3 and appendix A.3) were presented for answering.
- **Day 7:** Reviewing the system. These questions were about reviewing the whole system. Questions from the System Usability Scale (see section 4.1 and appendix A.1) were asked.

All diaries and questions can be found in appendix B.

Interview To end the evaluation an interview was held with each participant. Subjects of this interview were: their impression of the system: did it meet their expectations, etc. More in-depth questions were asked about the functionality of the system, and the content and timing of the messages. Participants were also asked about their perception of the two devices they could use the system on. Together with the participant, the interviewer looked at the diaries giving the participant the opportunity to clarify his or her entries. After this, questions about their privacy were asked and whether they missed something in the system. Finally, their opinion of the system was asked again. The interview script can be found in appendix D.1.

5.2.4 Data-analysis

In this section we explain the analysis we did on the data. First of all we made an inventory of the answers of the diary questions. Transcriptions were made of the end interviews. For each topic we combined the answers given in the dairy and interview, and an overview of all the given answers was made. The Attrakdiff questionnaire was processed, and calculations were made for each of the four constructs (for more information about Attrakdiff see section 4.3 or appendix A.3). An image was made from these results. Finally, the system usability score was calculated based on the answers given on day 7 of the diary (see previous section for the diary or appendix B).

5.3 Results

In this section we will provide the results of the evaluation of the SmarcoS-diabetic system. We will give the results per category. The categories are: using the system, medication messages, activity messages, system information and functionality, privacy, and other. After this we will give the Attrakdiff results and the scores of the System Usability Scale.

Although we asked participants to answer all the questions in the diaries, not all participants did this. Therefore, it can occur that when stating the results only three (or any other number of participants) out of five participants answered.

When multiple participants gave the same answer we will use the notation (3/9) for '3 out of 9 participants'.

5.3.1 Using the system

We asked the participants whether they would like to use the system themselves if it was further developed. One participant (1/5) answered that he would like to use the system himself. Another participant (1/5) answered that she would personally not use the system. Other participants answered evasive, indicating that they are doubtful whether they would use it or use just a specific part of the system.

Would you like to use a system like this yourself?
"I think I would like it to get more insight, since I want to be more active." (this person could not get the activity meter to function, therefore this is not based on any experience)
"What I found difficult was the pill dispenser since I don't take my medication at a set time, but rather with every meal. Personally I would not use the system, but I can imagine that it would be very useful for others."
"I think that when I would forget my medication more regular, I would go to search for a system that would help me with that."
"Yes, absolutely"
"I would like to use the activity monitor, I mostly think it is fun. I know I am physical active enough, but it is fun to see it and to try to be even more active."

All participants (5/5) indicated that they thought such a system could help people live a more healthy and balanced life, and that the time the system took to use was fine (not too long). Three participants (3/5) indicated that the system was easy to use and explained itself. The other participants (2/5) did not mention something about the user friendliness.

Do you think that a system like this can help people to get a healthier lifestyle?
"Yes, I do think so. I think you get more insight with the system, if you don't take your medication on time or if you are not active enough, it will get you thinking" (this person could not get the activity meter to function, therefore this is not based on any experience)
"Yes, I do think so. Using the system you are more conscious with your physical activity than you would normally be."
"Yes, I do think so, receiving feedback is stimulating for people."
"Yes, if it was more extensive. With food tips and stuff."
"I think it would work for the stereotypical diabetic. I am curious how they react to motivation. For me it is stimulating to see how active I was during the day. I want to reach the 100% target, but I don't know if it works like that for everyone."

5.3.2 System information and functionality

Participants were asked to place smiley stickers with different parts of the system. All smileys to choose from can be found in table B.1 of appendix B.5, while an example of a page of a completed diary can be seen in figure 5.7. The results can be found in table 5.1 and table 5.2.

This was repeated, but this time asking the participants to place words with the different parts of the system (for all word to choose from see figure B.1 of appendix B.6, for an example of a page of a completed diary see figure 5.8). The results can be found in table 5.3. Each smiley sticker and word sticker was only available once for each participant.

Figure 5.7: Example of a page of a completed diary with smiley stickers

Figure 5.8: Example of a page of a completed diary with word stickers

Table 5.1: Smiley stickers chosen and comments given about the computer application (one row per participant)

Computer application		
Activity history	**Medication history**	**Messages**
		
	 “is not always correct”	 “I did not get all messages”
 “fun to know”	 “clear”	 “did not get any ”
		
“I don’t have time to see this regularly’		

Table 5.2: Smiley stickers chosen and comments given about the smartphone, pill dispenser and activity monitor (one row per participant)

Smartphone	**Pill dispenser**	**Activity monitor**
		
 “is not correct”	 “fine”	
		
	 “opens too easily!”	
 “relaxed, great app. Too bad the activity monitor does not synchronise wireless.”		

During the interview participants were asked if they had the impression that they were using one system, although it uses multiple devices. All participants (5/5) answered that they thought it was one system. We asked them what they thought the differences were between the different devices. One participant (1/5) said there were no differences. Two participants (2/5) preferred using the computer more. The first participant preferred the computer, because the computer displayed the correct information while the information on the smartphone was incorrect. The other participant preferred the computer application, because the computer application was larger and therefore clearer. One participant (1/5) had a clear preference for the smartphone, which he found clearer. He found the application on the computer less clear because he could not distinguish between images which you could click on and the images you could not click on. The

Table 5.3: Words associated with system parts

Activity monitor	useful innovative motivating challenging presentable
Docking	complicated easy to use easy to use useful distant (“wireless = close”)
Pill dispenser	confusing useful useful easy to use hard to use(“too large”)
Messages	motivating clearly structured clearly structured motivating
Smartphone application	nice presentable coherent clearly structured useful
Computer application	coherent integrated clearly structured
Computer application activity history	motivating motivating
Computer application medication history	coherent pleasant
Computer application message overview	pleasant nice

last participant (1/5) liked the computer better because it contained more information. However, it was not the SmarcoS application he referred to but the DirectLife website to synchronize the activity monitor. He did use the smartphone application more, because he always carried it with him.

5.3.3 Medication messages

We asked our participants when they would like to receive a reminder to take their medication. Two participants (2/5) indicated that they would like a reminder after the medication moment, while one participant (1/5) would like to have a reminder before the medication moment. One participant (1/5) answered that he liked it the way it was now (30 minutes after the medication moment), but suggested to add a reminder if the medication still was not taken an hour later.

The participants indicated the following moments as moments that they would like to receive a reminder to take their medication: any time, at home before going out, and during the meals of the day. One participant (1/5) indicated that she would rather not receive a reminder, and that she therefore could not give an indication about a good moment to receive such reminder.

Moments that participants would not like to receive any medication reminder were:

at work, during grocery shopping, when they are visiting someone, on Sunday, during a meeting or at night.

When would you NOT like to receive a reminder to take your medication?
"At work, because it only would be chaotic for me."
"During grocery shopping or if I am visiting someone, because I usually don't take my medication with me."
"During a meeting or at night"
"You can read the message in your own time. If it is not convenient for you at the moment you get the message you wait until you can read it. But I would not like it if I receive messages at night."
"I would not like to receive messages on Sunday. Sunday morning at church I think I would not appreciate it! And at night."
"It does not matter when I receive the message, I can read the message when I want it."

Participants would like to receive feedback on their medication intake in the evening (3/5) or at the end of the day (1/5).

We also asked the participants what they thought of the content of the medication messages. They indicated that they felt motivated when they received a message containing 'well done!'. The messages were seen as clear, friendly and polite. Others thought the reminders to be useful and not irritating.

When participants compared the way they normally take their medication (all have their medication at a specific spot, and two participants use a ritual to not forget to take their medication) three participants (3/5) mention the timeframe in which the medication intake is correctly. Participants explained that medication is often taken with every meal, but mealtimes can vary due to other occupations. Since the medication time is set at a specific time this sometimes does not correspond. Participants told that they took medication out of the pill dispenser to be on time for the system, but subsequently forgot to take their medication. Others indicated being more conscious of the time and getting out of bed early during the weekend to be on time to take their medication and to avoid getting a reminder.

When we asked the participants about the timeframe around the medication moment, four participants (4/5) answered that it was fine this way, while one participant (1/5) would like to have a connection with meal times. When we asked them to indicate a good timeframe for themselves, three participants (3/5) answered the same timeframe as used, one participant (1/5) would like to receive a reminder before the medication moment, and one participant (1/5) would like to have a reminder 20 minutes after each meal.

5.3.4 Activity messages

One participant did not have a working activity monitor.

Participants indicated that they would like to receive a reminder to be more active in the morning (2/5), around lunch time (2/5), when they are at home (2/5), or if they are sitting still for too long (1/5). Moments when they would not like it to get a reminder to be more active were: at work, during grocery shopping, after 20:00, after 23:00, and during a meeting.

A feedback message about the physical activity done, participants would like to receive during dinnertime (3/5) or in the evening (1/5). An overview of the physical activity participants would like to receive at home (2/5), at the end of the afternoon (1/5), during lunch or dinnertime (1/5), and one participant (1/5) indicated always.

They said that they would not like to receive an overview of their activity when they are a sleep, on Sundays, after 23:00, one participant said it did not matter when, while the last participant said the same, but added as long as it's not a full screen message and you can read it later.

Two participants (2/5) mention that they would find it a huge improvement if the activity monitor would synchronize automatically.

"Too bad the activity monitor does not synchronise wireless."

"I would like the activity monitor to be more direct. If I forget to take it with me, I now get a message the next day which I think is too late."

5.3.5 Privacy

Participants were asked whether they had any objection to the information that is gathered by the system. All participants indicated that they did not have any objection to this. We asked participants whether they would have any objection if location was added to the gathered information (to be used to time messages better). Three participants (3/5) answered they still did not have any objection. While two participants (2/5) said they would not have any objection if location was not used standard and you could turn it off yourself.

5.3.6 User experience

User experience was measured asking the Attrakdiff questionnaire of Hassenzahl (for more information about Attrakdiff see section 4.3 or appendix A.3). One participant did not answer all items of this questionnaire; therefore results of this questionnaire are based on four participants. The results can be found in figure 5.9. The colours at the side indicate four aspects you can measure with those word-pairs [34, 35, 36]. The word-pairs at the yellow beam are used to measure pragmatic quality (this describes the usability of a product and indicates how successfully users are in achieving their goals using the product), the word-pairs at the green beam are used to measure hedonic quality – identity (this indicates to what extent the product allows the user to identify with it), the word-pairs at the blue beam are used to measure hedonic quality – stimulation (this indicates to what extent the product can support the need to develop and move forward), and the word-pairs at the red beam are used to measure attractiveness (this describes the global value of the product based on the quality perception). In this graph the orange line is the mean value that was given, while the greyish area indicates the standard deviation.

When we look at figure 5.9 we see that in general, word pairs score above average (average being between -1 and 1) or at least at the high end of the average. Words pairs that stand out negatively are: technical–human, separates me from people–brings me closer to people, cautious–bold, and undermining–challenging. The reliability of each dimension is not good enough, apart from the pragmatic quality, since we defined a construct to be reliable if Cronbach's alpha > 0.6 due to the small group. Removing a word pair from each dimension increases the reliability as following: removing technical–human from the pragmatic quality dimension increases reliability to Cronbach's alpha $= 0.88$, removing tacky–stylish from the hedonic quality - identity dimension increases reliability to Cronbach's alpha $= 0.65$ (which makes this dimension reliable), removing cautious–bold from the hedonic quality - stimulation dimension increases reliability to Cronbach's alpha $= 0.05$ (still not reliable), and removing bad–good from the attractiveness dimension increases reliability to Cronbach's alpha to 0.46 (still not reliable).

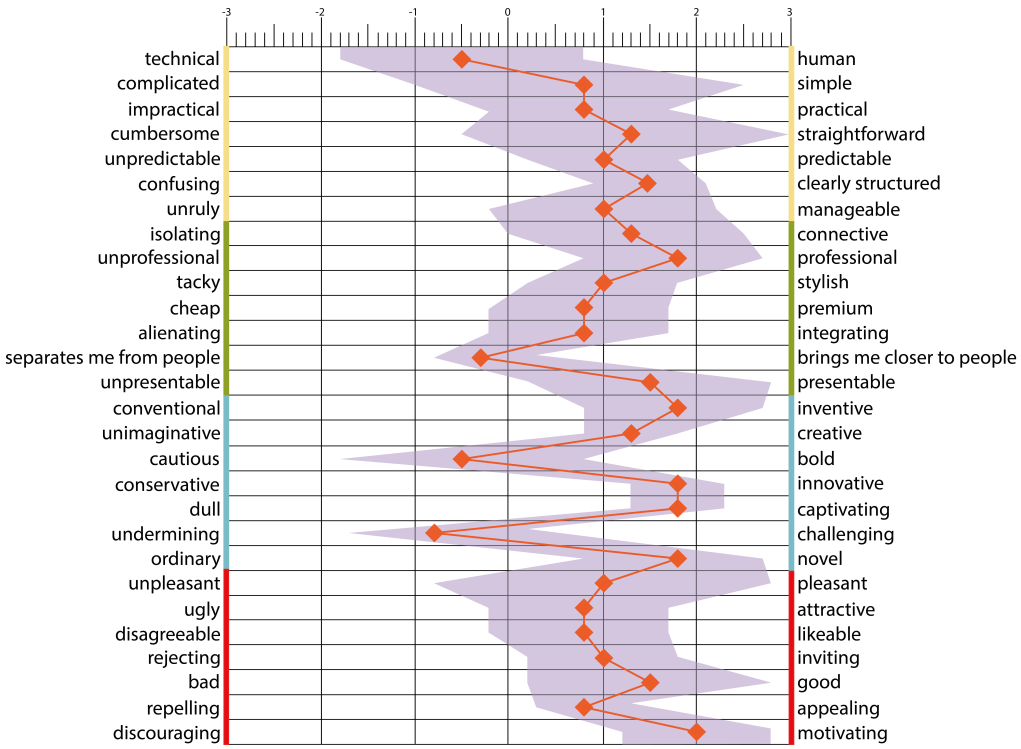


Figure 5.9: Attrakdiff results
pragmatic quality (yellow beam) Cronbach's alpha = 0.73,
hedonic quality - identity (green beam) Cronbach's alpha = 0.47,
hedonic quality - stimulation (blue beam) Cronbach's alpha = -7.88,
attractiveness (red beam) Cronbach's alpha = -0.02

The average values of each dimension are shown in figure 5.10. We already saw that *only* the first dimension is reliable when taking all word pairs into account; therefore we should be careful interpreting these average values. Averages for hedonic quality - stimulation and attractiveness lie above average, the others lie at the high end of the average region.

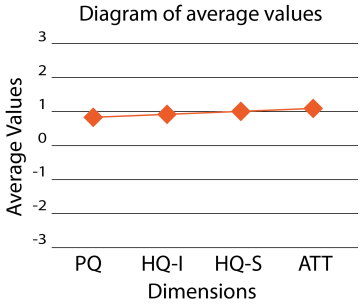


Figure 5.10: Attrakdiff results averages for each construct

5.3.7 Usability

We asked our participants to fill out the System Usability Scale (for more information about the system usability scale see section 4.1 or appendix A.1). The results can be found in table 5.4. The System Usability Scale is reliable with Cronbach's alpha = 0.84. System usability scores can range from 0 till 100. When we look at the scores given to the system we see that the lowest score is 55 while the highest score is 92.5. The mean score is 76.

Table 5.4: Usability Scores (5-point scale (1=strongly agree, 5=strongly disagree))

Statement	**Participants scores**					**Mean score**
I think that I would like to use this system frequently	5	1	3	5	3	3.4
I found the system unnecessarily complex	1	3	4	1	1	2
I thought the system was easy to use	5	3	5	5	4	4.4
I think that I would need the support of a technical person to be able to use this system	1	2	3	1	1	1.6
I found the various functions in this system were well integrated	5	4	3	3	4	3.8
I thought there was too much inconsistency in this system	1	2	3	1	2	1.8
I would imagine that most people would learn to use this system very quickly	5	3	4	5	4	4.2
I found the system very cumbersome to use	1	2	5	1	1	2
I felt very confident using the system	3	3	3	3	4	3.2
I needed to learn a lot of things before I could get going with this system	2	1	1	1	1	1.2
System Usability Score	92.5	60	55	90	82.5	76

5.3.8 Other

Other remarks were made about the medication moments. Participants would have liked to be more flexible, and adjust medication moments per day of the week, or specific days (2/5).

One participant would like to have messages more specific to the situation. The example given by the participant was: "Do not only say you don't exercise enough, but also give tips on what someone can do to be more active. Specific to the situation, if someone does not take his medication four days in a row, then something is wrong. Make sure you do something with that information, and give tips according to what is happening."

A participant would like to have an alarm in the pill dispenser, this way the participant gets an alarm signal when the medication moment occurs and a reminder after 30 minutes if the pill dispenser still is not opened.

One participant would like more functionality on the smartphone. He would like to be able to click on the medication clocks in the smartphone application. An example of what should happen was to show pictures of the pills that have to be taken. Also he would have like to see the time of the medication moment under the image of the clocks in the application.

A participant would have liked to also have some kind of glucose monitor in the application. Ideally he would have liked it to be automatically with a sensor in the cap of the insulin pen. He also would like to add a fruit and vegetable counter. And product information, so you can make the better choice when doing grocery shopping.

In this section we provided all results. Although all participants see the benefits of such a system and think it would work for others, only one participant indicated that he wants to use the system. We let participants associate words and smileys with different parts and functions of the system. We see a discrepancy in these results when we look at the computer application. The smileys (see table 5.1) are less positive than the words (see table 5.3). Medication reminders were correct, but the system leads to unwanted situations in which the pill is taken out of the pill dispenser before taking it. Therefore,

participants forgot the pill. Two participants (2/5) mention that they would find it a huge improvement if the activity monitor would synchronize automatically. The data that is gathered is not seen at too personal. The results of the Attrakdiff questionnaire lie above average (average being between -1 and 1) or at least at the high end of the average. Finally, system usability scores vary among participants. In the next section we will discuss the results in more detail.

5.4 Discussion SmarcoS-diabetic

In this section we will discuss the results of the SmarcoS-diabetic evaluation. We will take a closer look at the results and point out some interesting findings. We will look at contradictions in the results, and discuss how they may be explained. Finally, we will discuss how these results give us more insight in behaviour changing support systems like the SmarcoS system.

Medication and pill dispenser Medication reminders were mostly seen as useful, the messages were perceived as clear, useful, friendly, polite and not irritating. Most participants thought the time-frame in which to take the medication was fine the way it was. However, participants were less positive about the timing of the medication moment itself. Some participants would like to get a reminder to take their medication before the medication moment, others thought it was fine the way it was, while one participant suggested a second reminder if the medication still was not taken an hour later. The problems with the medication moment itself also explain why we got different answers to the same question in the diaries and during the interview. Medication is mostly taken with every meal. Therefore, the system would benefit from being able to recognize this situation instead of using a set medication time. A participant told in the interview that she got out of bed early in the weekend just to open the pill dispenser so she would not get a message that she was late. Another example was a participant who went to a festival during the weekend. The timing was normally set on, for example, 18:00, during dinnertime. However, because of the festival, he would eat a bit later than normal. The system does not take this into account for the timing of the medication message. If he waited to take his medication he would be 'late' according to the system. Because this participant did not want to be late taking his medication according to the system, he took his medication out of the pill dispenser and unfortunately forgot to take it later. This shows that such a system should be more adaptable in order to work properly and definitely should not lead to more mistakes when medication is concerned. Secondly, timing of the medication would benefit from being easily adjustable by the user himself.

Not all participants agreed on what they thought about the pill dispenser. This is because one participant thought the pill dispenser opened too easily, and another participant thought the pill dispenser was too large (which he found impractical). Negative words associated with the pill dispenser included confusing and hard to use (the latter was the participant who thought the dispenser was too large). The tray in the pill dispenser contained two or four compartments; this was impractical for one of the participants since she not always has four medication moments, but sometimes three. She would have liked the opportunity to have a tray with three compartments. Another participant would like to have an alarm in the pill dispenser. This way the participant gets an alarm when the medication moment occurs and a reminder on the applications after 30 minutes if the pill dispenser still is not opened. These remarks illustrate that when we want a system consisting of multiple devices to work, we have to make sure they work well individually, as well as complement each other. Each element needs to have a purpose, be user-friendly, useful for a range of people, but also contribute to the whole system.

Activity During the evaluation rules were triggered but messages were not sent due to bugs. Two messages were sent based on the data by employees of Evalan. Most participants only gave information about the timing of the activity messages (feedback in the form of reminders or an overview). They want to receive feedback at suitable moments, which they consider to be in the morning, around lunch time or for example if they are inactive for too long (they did not specify 'too long'). When messages are sent, whether these are medication messages or activity messages the content of the message should be relevant and situation specific. For example: if someone always takes his medication too late, the message should not only contain a reminder, but also tips on how to take medication on time. The same goes for activity. If someone is not physically active enough during the day, the message should not only contain 'you should be more active', but also contain tips on how to be more active. A participant said: "Do not only say you don't exercise enough, but also give tips on what someone can do to be more active. Specific to the situation, if someone does not take his medication four days in a row, then something is wrong. Make sure you do something with that information, and give tips according to what is happening." This illustrates that the system would benefit from being 'smarter'. It should be able to recognize more situations, and give more specific feedback including tips.

Docking Although docking the activity monitor does not take a lot of time and effort, docking is seen by participants as a problem. They would like the activity monitor to synchronize automatically. This eliminates the need of docking, and improves the feedback because your activity data is always known and up-to-date. Feedback can now only be given afterwards. This was seen by the participants as unsatisfactory. Having automatic synchronization also would be beneficial for recognizing situations and giving more appropriate feedback.

Word and smiley association We let participants associate words and smileys with different parts and functions of the system. When we look at the results of this for the computer application we see a discrepancy. The smileys (see table 5.1) are less positive than the words (see table 5.3) associated with the computer application. The smileys associated with the computer application contain a lot more frustrating and not satisfied smileys compared to the words associated with the computer application. Of course this is influenced by the limited assortment of smileys and words, but the difference is clear. The words associated with the application seem to be more focused on functionality, while the smileys are more focused on an overall feeling. Secondly, we would like to argue that the smileys used in this evaluation are not the best choice, because different explanations can be given about the same smiley. This is illustrated with the comments that accompany the smileys (see table 5.1). Thirdly, the words participants could choose from are displayed in different fonts. This will probably influence the perception of the words. How this has affected the results is unknown.

When we look at the words and smileys associated with the smartphone application they more or less show the same. Some smileys are a little less positive than the words associated with the application, but the associations are more positive and similar when we compare the answers of the smartphone application with the computer application. Smileys and words were quite positive about the smartphone application.

Application preferences The smartphone application was preferred over the computer application by some participants because it was always accessible and participants always had their telephone with them. Also, one participant found it difficult to distinguish between parts of the computer application you could click on to navigate further, and the parts you could not click on. This is understandable because they look the same.

The computer application therefore would benefit from redesigning and providing some clues to where you can and cannot click on. Participants who used their telephone less preferred the computer application. Reasons for preferring the computer application were more information and larger (which made it easier to read). Accessibility on multiple devices was one of the central points of SmarcoS. By knowing preferences of the user, and the reasoning behind the preference the system can be improved to better suit all users.

Data gathering and privacy None of the participants had any objection to the data that is gather by the system. Because as they said "It is for your own benefit". They also considered the data that was gathered not too personal. When we asked participants whether they would have any objection if location was added to the gathered information, most (3/5) participants still had no objection. However, two participants would only have no objection if it was not used by default, and could be turned off. This was mainly because, as they explained, not everyone needs it or wants it. This is a really important point. It illustrates that people are willing to give up some of their privacy if they see the need for it or benefit enough from it. Two participants (2/5) do not feel like they will benefit enough from it that they are willing to give up some privacy. However, they do see that others might benefit enough from it that they might be willing to give up some privacy. This also shows a boundary concerning private information. Location is seen as more personal than activity data, or medication intake.

System use We asked participants whether they would like to use the system themselves, only one participant was clearly positive and answered "absolutely". One participant gave a clear no to this question. Others would like to get more insight in their activity, but they do not give a real answer to the question. While all participants were relatively positive about the system, it is not clear whether they would like to use the system themselves. However, when asking them whether they thought the system would be helpful for others living a more healthy life, all participant answered yes. Part of the explanation of this difference between wanting to use the system yourselves and thinking it might be useful for others may be found in the fact that all participants were diagnosed at least 1.5 year ago with diabetics or even longer. Also all participants had a clear routine in taking their medication, and participants might have a different view of themselves in comparison with the 'other' people. One participant answered: "I think it would work for the stereotypical diabetic". And with stereotypical diabetic he gave the description of an inactive, slightly overweighed person with bad eating habits. Finally, the difference between using the system yourselves, and thinking it might be beneficial for others might be explained by missing functionality. Examples of functionality that should be added according to participants were glucose monitoring, pictures of the medication that has to be taken, the time to take your medication under the clocks of the medication time, a fruit and vegetable counter, and product information, so you can make the better choice when doing grocery shopping.

User Experience When we look at the results of the Attrakdiff questionnaire, we see that most word pairs score quite positive. The average of each construct (figure 5.10) lies around or above average (the average region lies between minus one and one). The hedonic quality - stimulation and attractiveness even score slightly above average).

First we will look at the pragmatic quality (yellow beam in figure 5.9). We see scores just below one, which is about average. Participants see the system as being technical. This may be due to the fact that it uses multiple devices, not only for measurement, but also for feedback. Whether this is a problem is not clear, because this is not something that came up in the diaries or during the interview. What did show in the diaries

and during the interview is that the system could be even more simple and practical. Comments given during the interview that the pill dispenser opened too easily, and the activity monitor should synchronize automatically will have contributed to these scores.

When we look at the hedonic quality - identity (green beam in figure 5.9) we see that participants feel like the system separates them from people. This might be explained by the written remark made by one of the participants concerning docking the activity monitor with word association (see table 5.3): "distant ("wireless = close")". The word pairs unprofessional–professional and unpresentable–presentable score quite high, indicating that participants find the system more professional and presentable than unprofessional and unpresentable. Other word pairs just score around average leaving room for improvement so participants can identify themselves better with the system.

When we look at the hedonic quality - stimulation (blue beam in figure 5.9) we see a very variable scoring. Some word pairs score high, others score low. The low scoring word pairs are: undermining–challenging and cautious–bold. This is possibly due to different interpretations of some scales, and different required results. For example, the system should help the user, which might mean that it should not be too challenging. Because the system is helping them to take their medication on time and be active enough, and therefore helps them to be healthier, it can be seen as a cautious system. It helps to prevent problems. When we do not take these word pairs into consideration hedonic quality - stimulation score well above average indicating that system can support the need to develop and move forward.

When we look at the attractiveness (red beam in figure 5.9) we see that all word pairs were scored relatively similar. Participants find the system more motivating. This is good since the system should be motivating. Of course it is always good to aim for improvement, but concerning attractiveness it is quite good already.

System Usability Scale We already discussed in section 4.1 what makes a good system usability score (see figure 4.1). We established that a system usability score of 70 or higher is good, while we aim for a score in the high 70s or even higher. When we look at the scores given by each participant the system usability score varies between 55 and 92.5. Two participants (2/5) score the system below a good system usability score. All other participants (3/5) score the system well above the good score. When we take the average of these scores the overall score of system is 76. This is above the good score. However, it is just above good; therefore there is still a lot of room for improvement. It also would be better if all participants had scored the system above 70. When we look at the individual scores of the items it stands out that for example "I found the system very cumbersome to use" was rated very different by the participants who scored the system the lowest in comparison to the participants who scored it best. The same goes for "I found the system unnecessarily complex". This indicates that improvement should be in this area.

6 Evaluation SmarcoS-office worker

In this chapter we will discuss the evaluation of the SmarcoS-office worker versions of the system (SmarcoS-office worker-ECA and SmarcoS-office worker-text). These versions are based on the first use case in which office workers are the target group. We will first discuss the two office worker versions of the SmarcoS system. After this we will discuss the methodology. In the methodology section (section 6.2) we will give most attention to the construction of the questionnaires. We will end by presenting the results and discussing them.

6.1 SmarcoS-office worker (SO)

The SmarcoS-office worker versions of the SmarcoS system focus on the first use case in which office workers are the target group. The two versions that are evaluated only differ in output modality. We will call them SmarcoS-office worker-text (SOT) and SmarcoS-office worker-ECA (SOECA). These versions differ from the SD version at main points. First of all both versions (SOT and SOECA) do *not* use the pill dispenser. Secondly, they *only* use the smartphone application. Thirdly, these versions have a different set of rules (rules are the same for both office worker version), and finally, the way the activity goal is set differs. We will now discuss all the last two points.

Rules Both versions (SOT and SOECA) use the same set of rules providing feedback based on the activity data. They also use the same set of texts that are sent to the smartphone application. Rules can be divided into categories: time-triggered and docking-triggered. We will now shortly discuss all rules, for each rule we will give an example sentence that is send to the user.

Time-triggered:

- **RequestToDockNoDocking:** When it is after 18:00h and the user has not docked his AM, send a "docking request no docking" between 18:00–24:00.
  "Je hebt vandaag nog niet je Activity Monitor gedockt, probeer deze nog aan te sluiten."
  There are fifteen variants on this sentence.

- **RequestToDockDockingTooLongAgo:** When it is after 18:00h and the last time the user has docked his AM is more than two hours ago, send a "docking request too long ago" between 18:00–24:00.
  " 3 uur geleden heb je je Activity Monitor voor het laatst gedockt. Het is nu weer tijd om dat te doen."
  There are 23 variants on this sentence.

- **WeekOverviewBelow80:** When it is a Monday, it is after 21:00h, and the average week percentage of the user is below 80% send a "week overview below 80 message" between 21:00–24:00.
  "Afgelopen week heb je 70% van je doel bereikt. Probeer komende week harder aan de slag te gaan."
  There are 20 variants on this sentence.

- **WeekOverview 80-99:** When it is a Monday, it is after 21:00h, and the average week percentage of the user is between 80% and 100% send a "week overview below almost message" between 21:00–24:00.
  "Afgelopen week heb je bijna je doel bereikt. Met een klein beetje meer inspanning haal je volgende week je doel."
  There are ten variants on this sentence.

- **WeekOverview 100>:** When it is a Monday, it is after 21:00h, and the average week percentage of the user is more than 100% send a "week overview 100 message" between 21:00–24:00.
  "Gefeliciteerd Jan, deze week heb je 118% van je doel bereikt. Probeer dit de komende week ook vol te houden."
  There are eighteen variants on this sentence.

In practice messages were send at the time of the check (18:00 or 21:00 depending on the message), or at the minute when the last message was sent two hours ago.

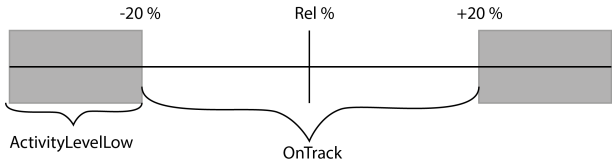


Figure 6.1: When is a user on track? When the current percentage of the user is within 20 percent of the relative percentage (Rel = (100 %/((Now - 7 hours)/(23 - 7))))

Docking-triggered:

- **OnTrack:** When the user docks his AM, and he is 'on track' ('on track' means current activity percentage is within 20 percent of the following percentage: Rel = (100 %/((Now - 7 hours)/(23 - 7))) see figure 6.1, send a "on track message".
  "Je bent goed op weg om vandaag je doel te halen. Ga zo door."
  There are 20 variants on this sentence.

- **ActivityLevelLow:** When the user docks his AM, and his activity level is 'low', ('low' means current activity percentage is lower than the following percentage: Rel = (100 %/((Now - 7 hours)/(23 - 7))) minus 20%, see figure 6.1, send a "activity level low message".
  "Je bent vandaag nog niet zo actief geweest, misschien moet je meer gaan bewegen. Heb je je Activity Monitor wel altijd bij je?"
  There are 20 variants on this sentence.

- **GoalReached:** When the user docks his AM and his current activity percentage is between 100% and 120% , send a "activity level reached message".
  "Gefeliciteerd, je hebt je doel vandaag gehaald. Probeer dit vol te houden."
  There are 45 variants on this sentence.

- **GoalReachedHigh:** When the user docks his AM and his current activity percentage is higher than 120%, send a "activity level reached high message" .
  "Heel goed gedaan, je hebt 134% van je doel gehaald. Ga zo door."
  There are 21 variants on this sentence.

There is one difference between the two office worker versions of the SmarcoS system. The SOT version, displays the feedback via text, while the SOECA version displays the feedback by letting an ECA speak the text. For information about the ECA used in the

application and how this ECA is realised on a mobile phone we refer to section 3.2. In both versions participants still can see their activity history.

Smartphone application As can be seen in figure 6.2 the smartphone application has two icons each representing a function. We will discuss each of these icons, combined with their functionality. First of all: the left icon. When selecting the left icon, the user can see his activity overview since the end of the assessment week (start evaluation). The icon itself shows the last known activity data progress in percentage (in this case 103%) towards the activity goal of that day. Or, if the user has not yet docked his activity monitor, it states please dock.

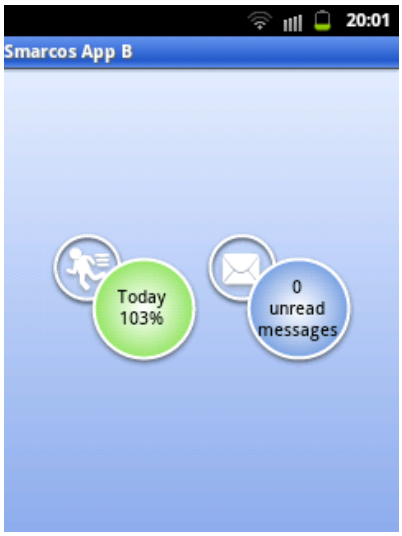


Figure 6.2: Smartphone Application versions SO

The other icon (at the right in figure 6.2) shows how many unread messages the user has (in this case there are no unread messages). By selecting this icon the user is shown his inbox of messages. When selecting one of the messages the user can read the whole contents of the message. Like an e-mail inbox, a subject of the message is shown as well as the time the message is send. Opening a message the SOT version just shows the content of the message in text. In the SOECA version opening a message starts the ECA. A screenshot of the ECA can be found in figure 6.3. The message is then read out loud by the ECA, some limited facial expression is shown depending on the content of the message (a reminder for being more active, the ECA will have a more sad expression, a positive message, the ECA will look happy). The ECA is able to blink her eyes, and move her mouth when speaking. For more information on the ECA we refer to section 3.2.

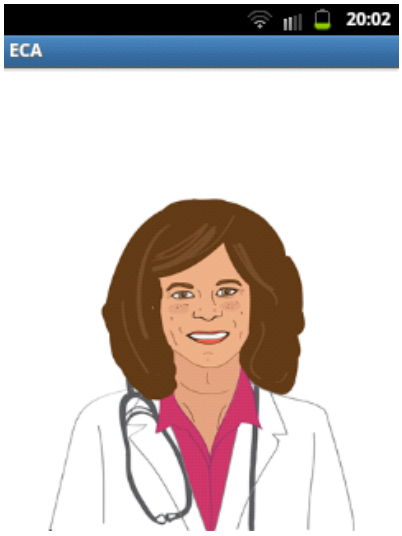


Figure 6.3: Screenshot of the used ECA

Setting the activity goal In the evaluation of the SO systems an assessment week is used to first determine the normal level of activity for each participant. This was chosen because in these versions motivation to be more active is really important. Without a correct goal participants might receive messages that would be correct for their activity goal, but might be incorrect in motivating them to be more active. Again, a goal that is too high will lead to getting reminders to be more active, which will frustrate users instead of motivate them to be more active. When the activity goal is too low, it leads to easily receiving messages that you have reached your activity goal. The latter makes it impossible to see if the system indeed can motivate the user to be more active.

Generation of behaviour change support system The SOT of the SmarcoS system belongs to the third generation of behaviour change support systems, since it runs on a smartphone and gives tailored feedback on the situation. When we look at the SOECA version of the system, we can argue that this version belongs to generation 3.5. Since the addition of an ECA can be seen as a next step in the evaluation of behaviour change support systems.

6.2 Methodology

In this section we will first provide information about the participants, after which we will discuss the procedure. Finally, we will tell something about the materials and shortly explain the data-analysis.

6.2.1 Participants

Table 6.1: Experience of participants with: smartphones, behaviour change support systems, coaching systems and ECAs)

Experience with:	ECA participants	Text participants	All participants
smartphone	(M=8.5, SD=1.3)	(M=7.8, SD=0.8)	(M=8.1, SD=1.1)
behaviour change support systems	(M=2.0, SD=0.8)	(M=4.8, SD=3.0)	(M=3.6, SD=2.7)
coaching systems	(M=2.3, SD=1.0)	(M=4.8, SD=2.5)	(M=3.7, SD=2.3)
ECAs	(M=2.8, SD=1.7)	(M=3.8, SD=2.4)	(M=3.3, SD=2.4)

The participants in this study were nine office workers. They were divided into two groups depending on their smartphone (the ECA version of the system cannot run on all android versions). Participants were between 20 and 57 years old (ECA group: M=32.3 year, SD=16.8, Text group: M=32.2 year, SD=11.8, Combined: M=32.2 year, SD=13.3). We had five male participants and four female participants (ECA group: 4 male, Text group: 1 male, 4 female). All participants own a smartphone and use it daily (one has a blackberry on which the application does not run, for her we provided another smartphone). Participants indicated that they sit behind a desk for between 6 and 10 hours a day (ECA group: M=7.3 hours, SD=1.4 hours, Text group: M=7.8 hours, SD=1.3 hours, Combined: M=7.6 hours, SD=1.3 hours). Experience of participants with smartphones, behaviour change support systems, coaching systems and ECAs can be found in table 6.1.

Five participants (5/9) answered that they found themselves active enough, while three participants (3/9) answered that they were not active enough, and one participant answered it differed per day. One participant answered that she was less than 30 minutes active a day, while three participants (3/9) answered that they were more than 30 minutes active a day. The remaining five participants (5/9) answered that they were just 30 minutes active a day. Sports that were done by participants can be found in table 6.2

Reasons that were given to be more active can be found in table 6.3 (the reasons found in table 6.3 were the reasons participants could choose from, they also could add a reason themselves with none did).

6.2.2 Procedure

The duration of the evaluations was two weeks. At the beginning of the test, each participant was visited at home/work. Each participant received the activity monitor.

Table 6.2: Sports that were done by participants including frequency (one row per participant)

Group	Sport	Frequency
ECA group	fitness	1x a week
	running	2/3x a week
	cycling	to school not every day
	knotsbal	1x a week
	crossfit	3x a week
Text group	running	1/2x a week
	canoe	3x a week
	tennis	2x a week
	running	3x a week
	running	1/2x a month
	spinning	2x a week
	ballroom dancing	3x a week

Table 6.3: Reasons given to be more active, number of times answered per reason)

Reasons to be more active:	ECA group	Text group	Combined
stressreduction	1	3	4
increasing energy	1	2	3
for fun	1	1	2
seaking a challence	1	0	1
being admired	0	0	0
social activity	0	1	1
increasing the odds of winning in sports	0	1	1
doctor's advice	2	2	4
disease prevention	0	1	1
improving general health	3	3	6
weight control	3	1	4
to look better	1	1	2
increasing fitness	3	3	6
increasing flexibility	1	1	2

Participants were required to own a smartphone with android version 2.3 or higher (as already said before, one participant was provided with a suitable smartphone by us).

During this visit the participants were explained the goal of this study and the procedure. A consent form was given in combination with an information sheet. After reading this and making sure there were no questions left, participants signed the consent form. The information sheet also contained information for the installation of the required DirectLife software. E-mail addresses of all participants were asked to provide links to questionnaires and the smartphone applications.

It was explained that the first week was an assessment week, to establish their normal activity level in order to tailor the system to this. After the assessment week participants used the system for a week. Participants were asked to dock their activity monitor regularly during the assessment week and the test week. During the assessment week no feedback was given by the system.

During the test week, feedback was given by the system, and participants were asked to fill in a questionnaire each day starting at day four.

Shortly after the test week, an interview took place with each participant. For a description of this interview we refer to the next section.

6.2.3 Materials

In this section we will provide information about the materials used in this evaluation. We will discuss the questionnaires the participants were asked to fill out, and the interview that was held afterwards.

Questionnaires During the evaluation participants were asked to fill out five questionnaires. The questionnaires have the following title and will be asked in this order:

- 'Welcome' questionnaire
- 'Timing of the messages' questionnaire
- 'Content of the messages' questionnaire
- 'Coaching' questionnaire
- 'End' questionnaire

The questionnaires are based on the questionnaires discussed in the chapter 4. Before making an overview of these questionnaires as is shown in chapter 4, an inventory was made of the questions we wanted to be able to answer, or at least get more insight in. We will discuss the questions we want to answer and how we try to answer these questions per category/subject. We will explain where the questions used in the questionnaire originate from, and why we choose these questions in the questionnaire.

Demographics and experience In the first questionnaire questions are asked about the demographics of participants, this way we can characterize our participants. Age and gender are asked. We ask participants for which they normally use their smartphone to get an indication how familiar they are with it.

To get a better impression on how active our participants are we ask how many hours a day they spend sitting behind a desk (this also will give confirmation that they belong to our target group of office workers). Participants are asked if they think they are active enough and if they are active for an average of 30 minutes a day. We ask participants if they do any sports, and if so how often. A list of reasons to be more active is presented and participants are asked to indicate which reasons are legitimate for them at this moment to become more active.

Finally, participants are asked about their experience concerning smartphones, behaviour change support systems, coaching system, ECAs and mobile internet. These questions will be used to characterize our participants and might be used to explain results.

Expectations We want to know if the system meets the expectations of the user. This might explain why participants do or do not want to use the system. For this we added an open-ended question to the 'Welcome' questionnaire (appendix C.1). In this question we ask participants to indicate what they expect from the system. Suggestions are given: What will the system do?, Which benefits does the system have for me?.

We also wanted to know beforehand expectations of the user on the following constructs of technology acceptance: trust, anxiety, attitude, and performance expectance. These will be measured in twelve items also used in another part of the questionnaires and can be found in appendix C.1. More information about these constructs and why they are chosen can be found in 'technology acceptance'. We will also give some attention to expectations in the concluding interview.

Activity message There are several questions we want to have answered concerning the messages. We want to know what participants think about the timing of the messages and what they think about the content of the messages. Questions that we will try to get a partial answer on in the questionnaires are: Does the system send useful messages? Does the system send messages at the right timing? Are the received messages divers enough? We will first discus questions about the timing.

It might be that the timing of the messages was terrible. There are several open-end questions that will be asked about timing of the messages, because we also want to know the reasons behind their answers. The questions that will be asked concerning the timing of the messages include: During the last days when did you receive messages? What did you think of the timing of these messages? When would you like to receive a reminder to be more active? When would you *not* like to receive a reminder to be more active? Depending on the answers that are given to these questions, the interview will be more elaborate on this subject. Questions about the timing of the messages will be asked in the second questionnaire 'Timing of the messages' (see appendix C.2).

The questions that will be asked concerning the content of the messages include: What do you think about the content of the messages? Which messages did you find pleasant to receive? Which messages did you find UNpleasant to receive? Do you think the content of the messages was correct? Did you find the messages motivation? Which kind of messages would you like to receive as well? We also want to know if users think the messages are divers enough. To answer this, the answers of the questionnaire will be checked. When participants did not state anything about the diversity of the messages a question about this will be asked in the end interview. Questions about the content of the messages will be asked in the third questionnaire 'Content of the messages' (see appendix C.3).

Technology acceptance Technology acceptance also is a factor we looked at (see section 4.4). Users should accept the technology, and therefore the system. Without accepting the system a user will not use it, no matter how good the product is or how useful. We want to know the acceptance of both office worker versions, so we can compare the two versions on acceptance. There are many questionnaires about ECA's and their personality. However, there is not a good questionnaire concerning the acceptance of ECA's by users. Robots come most close to ECA's when we look at the acceptance of them, especially since robots have improving social abilities, and more functionality (section 4.5). Therefore, we also looked at robot acceptance. It would be convenient to have the same questions for both versions of the system. For one, this makes it easier to compare the two versions, but it also requires less work making the questionnaires and no mistakes can be made giving the wrong version of the questionnaire to the participant.

We looked at three technology acceptance models: TAM (section 4.4.1, appendix A.4.1), TAM2 (section 4.4.1, appendix A.4.2) and UTAUT (section 4.4.3, appendix A.4.3), and three robot acceptance questionnaires (section 4.5): Heerink (appendix A.5.1), the Almere model (appendix A.5.2) and GODSPEED (appendix A.5.3). The Almere model is an extension of the questionnaire of Heerink [41], and is based on UTAUT. Keeping this in mind it is a logical step to use UTAUT as guide to measure acceptance, because we can then have one questionnaire for both versions. We made a list of all measured constructs in all previously mentioned questionnaires, after which we identified the constructs we were most interested in (we did not use all constructs because we felt that this would become too much). We will now shortly discuss each construct we measured.

- **Anxiety** is measured in multiple questionnaires, UTAUT was used. Item 26 was omitted because it is impossible to lose information by hitting the wrong button in this system. For formulation we looked at the Almere model.

- **Trust** is only measured in the Almere model. We used the whole construct.
- **Attitude towards the system** is measured in multiple questionnaires, UTAUT was used. We rephrased the first item of the construct to 'good' (item 9) because having 'bad/good' in a statement is confusing.
- **Intention to use** is measured in multiple questionnaires, UTAUT was used. All three items were summarized into one statement, because three items take more time to answer while the essence is the same. We also had the possibility to elaborate on this subject during the end interview.
- **Facilitating conditions** is measured in multiple questionnaires, UTAUT was used. We omitted item 19 and 20 because this was a first user evaluation and this was not part of our interests (compatibility with other systems I use, and a specific person is available for assistance with system difficulties)
- **Perceived ease of use** is measured in TAM, TAM2 and the Almere model. In this case we used TAM as a guide for this construct, but also paid attention to formulation in the other questionnaires. We omitted the item 'It would be easy for me to become skilful at using the system', because the system does not need the user to become skilful in order for the user to use it.
- **Performance expectancy** is measured in multiple questionnaires, UTAUT was used. We omitted the last item of the construct (item 4), because this was not applicable for the system (using the system is unlikely to get you a raise).
- **Usefulness** is measured in multiple questionnaires; we merge all items in the questionnaires to three items. We used the Almere model as guide for formulation.
- **Adaptability** is only measured in the Almere model, therefore we used this.

In the questionnaires we each time combined 2 constructs into one question, while we randomized the order the items. We combined the items of adaptability and trust into one question in the second questionnaire 'Timing of the messages' (second to last question, see appendix C.2). We did the same with the items of anxiety and attitude which also can be found in the second questionnaire 'Timing of the messages' (last question, see appendix C.2). Facilitation conditions, and perceived ease of use were combined in the third questionnaire 'Content of the messages' (second to last question, see appendix C.3). Finally, performance expectance and usefulness were combined to one question in the third questionnaire 'Content of the messages' (last question, see appendix C.3).

As mentioned before we used some of the constructs to help formulate an expectancy of the system. These were trust, anxiety, attitude, and performance expectance. Answers to these constructs will be compared (expectancy vs. after using the system). We will also look if there is a difference in answering of these constructs between the two office worker versions of the system (ECA vs. text).

Coaching For a user to accept coaching that is given, several factors are important. Coaching is given in this system by giving feedback, and the timing and content of the messages will influence this. We want to know what the user thinks of the quality of the coaching the system gives, and if the coaching source (the system) is seen as reliable. This will help answer questions as: Does the user trust the system and the advice it gives?, Is the coaching source seen as reliable? Finally, this will also influence the attitude of users towards the system.

When we look at the system we see that it is a behaviour change support system, and that it provides coaching of the users. Based on this we also wanted to know

how the user experienced the quality of coaching. This is related to the timing of the messages and the content of them, but from another point of view. Philips has made an adaptation of the Coaching Behaviour Scale for Sport (CBS-S) to evaluate the coaching provided by its DirectLife system. CBS-S can be found in appendix A.7.1 and the adaption on this from Philips can be found in appendix A.7.2. To use the version of the questionnaire that Philips made is obvious, because the system builds on DirectLife. However, some adaptations were needed, because the system works with different kind of feedback some statements were not applicable (this applies to all removed items being item 1, 11, 12, 13, 14, 15, and 16 of the original questionnaire in appendix A.7.2), and therefore impossible to answer for this system. Fourteen questions of twenty-one of the version of Philips remained, which can be found in the fourth questionnaire 'Coaching' the first two questions (see appendix C.4).

Secondly, we look at source credibility. Source credibility is important in this system. When the user sees the system as a credible source, they are more likely to follow its advice and act based on the feedback that is given [64, 85]. A well-known measurement is the set of source credibility scales of McCroskey [63]. For more on this we refer to section 4.6. We chose to use the twelve item semantic differential scales, because reliability is higher for this scale than for the fifteen item version, and construct validity was shown. This version of the scale is used more [63], and because it has less items it also takes less time to fill out. Again important, because we do not want to overwhelm our participants by presenting them long questionnaires. The source credibility scale will be asked in the fourth questionnaire 'Coaching' the final question (see appendix C.4).

User experience We wanted to know how users would experience using the system. We chose a questionnaire to measure user experience over other methods such as: experience sampling method [21], day reconstruction method (DRM) [48], and diary methods [13], because it would be faster, less interruptive and less demanding of the participants than one of the other methods. AttrakDiff 2 is used to measure user experience. Reasoning for this was that the questionnaire is relatively fast and easy to fill out, while measuring pragmatic quality (PQ), hedonic quality (HQ) (two dimensions; identity and stimulation) and attractiveness. Also, reliability is shown for hedonic quality - stimulation (Cronbach's alpha 0.79 - 0.90), hedonic quality - identity (Cronbach's alpha 0.73 - 0.83) and pragmatic quality (Cronbach's alpha 0.83 - 0.85) [38]. This is shown in other studies as well [35]. Finally, analysis of the results is relatively fast. The Attrakdiff questionnaire will be asked in the last questionnaire 'End' questionnaire (the first 3 questions, see appendix C.5).

Usability Whether users want to use a system depends on multiple factors, usability is one of these factors. To answer what users think of the usability of the system we used the SUS questionnaire already discussed in section 4.1. Other methods involving usability testing can be: think aloud protocol [33, 76], remote usability testing and lab usability testing [87], and walkthroughs [49]. We chose the SUS questionnaire, because it will answer all questions we have concerning usability. We preferred a questionnaire above other methods because it is an uncontrolled experiment and not a lab study. Secondly, it requires less time from participants as well as researchers. This is important since we have very limited time to conduct the evaluation. We chose to include the SUS questionnaire above other questionnaires, because it is a short questionnaire that can be used for a wide range of technologies. SUS is used often and is preferred above other longer questionnaires, because other questionnaires do not provide a better insight in usability of a system [88, 6]. Also, because we want to answer many questions, questionnaires easily become too long. This is something that should be avoided in

order to have a high percentage of completed questionnaires. SUS is proven to be reliable and has a high face validity [6]. SUS will be asked in the last questionnaire 'End' questionnaire the final question (see appendix C.5).

Interview To end the evaluation an interview was help with each participant. Subject of this interview were the following:

- expectations (and if they were correct)
- general impression
- daily use
- coaching and feedback
- privacy
- other (diversity texts)

Part of the interview was standard; questions were added based on answers that were given by the participants in the questionnaires. Therefore, not all questions were similar for all participants. The standard interview script can be found in appendix D.2. Questions were skipped if participant had already answered them in a previous question.

6.2.4 Data-analysis

In this section we explain the analysis we did on the data. First of all we made an inventory of the answers of the online questionnaires. A transcription was made of the end interview. For each topic we combined the answers given in the online questionnaires and interview (per participant), and an overview of all the given answers was made (all participants). The coaching's questionnaire and source credibility questionnaire were processed. The Attrakdiff questionnaire was processed, and calculations were made for each of the four constructs (for more information about Attrakdiff see section 4.3 or appendix A.3). Figures were made from these results. The system usability score was calculated based on the answers of this part of the online questionnaire. All other constructs were processed as well. If applicable, reliability of the questionnaire was calculated. Finally, a comparison between the two versions (ECA vs. Text) was made.

6.3 Results

In this section we will provide the results of the evaluation of the SmarcoS-office worker systems. We will give the results per category. The categories are: expectations, using the system, system information and functionality, activity messages, technology acceptance, coaching, privacy, user experience, usability, and other.

Two participants from the SmarcoS-office worker-ECA group either did not receive any messages or no fully functioning messages on their smartphone. They reported this during the interview, after completing the whole evaluation. One participant did receive messages and saw the ECA, but his text-to-speech did not function. He could see facial expression. He did not discover that he could read the last send message also on the website. The other participant had a synchronization problem; he did not receive any message on his smartphone, but did discover he could read the last send message on the website. Because both participants did experience the system, although with limitations, we decided to include both participants in the study. In some parts of

the results those participants are not taken into account, this will be reported in these sections.

Constructs are measured on a seven-point scale, unless otherwise noted. Reliability is calculated in the form of Cronbach's alpha. Normally a Cronbach's alpha of 0.7 is seen as reliable. However, because of the small group(s) we will treat a Cronbach's alpha of 0.6 of higher as reliable.

The results comparing the two office worker versions (ECA vs. Text) will be given at (the end of) each category (if applicable). An independent T-Test was used for this for each questionnaire. When multiple participants gave the same answer we will use the notation (3/9) for '3 out of 9 participants'.

6.3.1 Expectations

Which expectations do you have about the system?
"I have no expectations" (3/9)
"Not useful, and I am afraid the system wants me to be more active at moments I cannot do something."
"Funny, but something I would quickly get tired off."
"Giving better insight in activities & good feedback."
"Creating awareness that I have to be more active."
"Be more active, do my exercises more often, be more aware of my posture."
"Give a plan for burning calories, measure the burned calories, and give reminders to be more active."

In the first questionnaire we asked participants about their expectations. This was done in two ways: by rating statements and in their own wording. Three participants (3/9) reported they had no expectations. Other participants mentioned expectations as creating awareness to be more active and getting more insight in their activity, as well as giving feedback and getting a plan to be more active. Two participants (2/9) expected that they either would not like the system because it would demand them to be more active at moments that did not suit them, or that they would get tired of it quickly.

Did your expectations turn out to be right?
N/A (3/9)
"I was afraid to get messages at night while I was asleep, but that did not happen."
"I quickly got bored by it, and became annoyed and irritated. This is not a system I would use."
"I got more insight in my activity, but the feedback was bad. The number of messages depends on the number of times you dock the activity monitor. I did not find the content of the messages specific enough."
"No, because you get messages at irregular intervals. And if you are really busy at work you forget to dock the activity monitor. If you do not dock the activity monitor you will not get any messages. And a lot of times I got a message when I just got home."
"No, I got all my messages at 18:00. I would like to get them during the day. Also messages are not concrete."
"The number of burned calories was shown. It did give a plan, but it is questionable whether it was a good plan. And giving reminders to be more active, not really."

We asked the participants if their expectations came true. Three participants (3/9) could not answer this question since they did not have any expectations. In general the system did give participants more insight in their activity, but did not meet expectations concerning the plan and giving reminders.

Table 6.4: Expectations:
The constructs measured are from top to bottom: T = trust, An = anxiety, Att = attitude and PE = performance expectancy. Constructs were measured on a seven point Likert-scale (1 = strongly disagree, 7 = strongly agree).
Underlined scores exceed expectations (the scoring after using the system is higher than the expectation).
Note, in the second construct a lower score is more positive for the systems.
* this construct is measured on a five point Likert-scale

Construct	Statement	ECA group		Text group		Combined	
		expectations	afterwards	expectations	afterwards	expectations	afterwards
T*	I would trust the advice the system gives me.	3.2	3.0	3.2	2.8	3.2	2.9
	I would follow the advice the system gives me.	3.0	2.5	2.8	2.4	2.8	2.4
An	I feel apprehensive about using the system.	1.8	1.5	1.7	1.4	1.7	1.4
	The system is somewhat intimidating to me.	2.5	2.0	2.3	2.0	2.3	2.0
	I hesitate to use the system for fear for making mistakes, or breaking something.	1.8	1.3	2.6	2.2	2.2	1.8
Att	I think it is a good idea to use the system.	4.0	3.3	3.9	3.4	3.9	3.3
	The system would make my life more interesting.	3.0	3.0	1.4	2.4	2.1	2.7
	Working with the system is fun.	3.5	2.8	3.6	3.8	3.6	3.3
	I like working with the system.	4.0	2.8	3.4	3.0	3.4	2.9
PE	I find the system useful to be more active.	3.5	2.5	3.1	2.8	3.1	2.7
	Using the system enables me to accomplish my goal to be more active.	2.8	3.3	2.9	1.8	2.9	2.4
	Using the system I am more active.	3.3	2.8	3.3	3.2	3.3	3.0

When we look at the ratings that were given to the statements (see table 6.4) we see that expectations of the system were higher than the actual experience. Constructs that exceeded expectations (scoring after use is higher than the scoring before use) are underlined in the afterwards scoring. In the second construct (An = anxiety) a lower score is better due to the formulation of the statements. This is the only construct for which all items exceed expectations.

When we compare the two groups who used the two different versions of the system, we find no difference between formulated expectations, or if their expectations came true. Comparing the answers of the measured constructs we do not find a significant difference between the two groups (not in complete constructs or individual items).

6.3.2 Using the system

When we asked participants whether they would like to use the system themselves; eight participants (8/9) answered 'No'. This answer illustrates this best: "I would not use a similar system, but I can imagine that if it worked differently it would work." One participant answered that she would only use the system to monitor her activity, but that she did not like the rest of the system. Subsequently, we asked participants how the system should be changed in order for them to want to use the system. Participants want more specific and different messages, tailored to the situation. Multiple participants said

they would like to omit the docking of the activity monitor. Also timing should be better, and in some cases more frequently. The following suggestions were made as result of this question.

What should be changed in the system in order for you to use it?
More specific information.
More different types of messages.
More different messages.
'Please dock your activity monitor' is not a useful message.
The content of the messages is standard. (2/9)
Add a meal plan or recipe suggestions.
Not having to dock the activity monitor. (4/9)
The information in the smartphone application is disappointing.
The activity pattern should also be shown on the smartphone application.
Timing of the messages at better moments (if you can do something with it).
More interactive (better suited for the exact situation).
Do not say in a message 'by more active' when someone just was active.
The timing and content of the messages should also be linked to the history and activity pattern of that person.
You should specify yourself on which points you want to be corrected.
There should be more moments in a day when you get feedback.
Add a social norm. Regarding others, preferably self-chosen groups.
Having an automatic plan is not pleasant (without adding your own preferences).

Five participants (5/9) thought such a system could help people live a more healthy and balanced life, but *only* if the users want to change themselves. They think that if a user does not want to change the system would not help them. One participant thought that the system would be mostly useful to help people to become more aware that they should be more active. Another participant said that she thought the system could help people live a more healthy and balanced life, but only if the system would be 'smarter'. Finally, a participant thought the system would help people, but only if the system took into account the specific situation of the user. Who is the user and what motivates the user? The system should use this information generating feedback.

There is no difference in the answers given by participants who used different versions.

6.3.3 System information and functionality

We asked participants what they thought of the time it takes to use the system. Seven participants (7/9) answered that this time was fine. It was also mentioned by seven participants (7/9) that the website used by the activity monitor was extremely slow (too slow). One participant reported that she probably did not use the website properly, due to it being too slow.

User-friendliness of the system was considered 'good' by four participants (4/9). Two participants (2/9) would have liked to have a 'read all' button on the smartphone application to improve user-friendliness. One participant answered he would have liked to see his activity pattern on the smartphone application. Two participants (2/9) thought installing all applications and adjusting all setting was too much work and bothersome (these participants belonged to the ECA group who had to install multiple applications and adjust multiple settings). Docking the activity monitor was not considered user-friendly by one participant. It was not considered user-friendly to have a percentage

when the user does not know what his goal is, according to one participant.

Participants reported that they used the following parts of the system the most:

What did you use the most?
Activity monitor (3/9)
Activity pattern during the day (website)(2/9)
Website
Percentage on the smartphone application
Percentage on the website when synchronizing the activity monitor (2/9)
The smartphone application

Participants reported the following errors:

Did you experience any errors?
The server was down for a couple of days.
I got a docking request when I already reached my target for that day.
The system does not work when running Windows 8.
The system did not register it when I went for a run on Thursday.
While giving feedback the system should have taken into account the amount of intensive and average activity.
I got a 'You are doing well' message at 8:00 while I only got dressed and drove my car to work.
Depending on how you wear your activity monitor it does not always register the activity correctly.
The activity monitor was not recognized when docking it.

When we asked our participants which kind of extra functionality they would like to have, five participants (5/9) answered that they would like it if they did not have to dock the activity monitor anymore. Two participants (2/9) would like it when they could also see the activity pattern during the day on the smartphone application. All given suggestions are:

What kind of functionality would you like to include?
I would like to get a message when I reach 50%.
Not having to dock the activity monitor. (5/9)
A faster website.
The activity pattern should also be shown on the smartphone application. (2/9)
Better suggestions to be more active.
More specific messages.
Be able to compare myself with others.

There is no difference in the answers given by participants who used different versions.

6.3.4 Activity Messages

In this section we will provide the results of the activity messages. We will first show the results about the number of messages, after which we will give the results concerning timing. Finally, we will give the results about content.

Number of messages When we asked participants about the number of message they got, two participants (2/9) thought there were not enough messages, while two participants (2/9) thought they got too many. Two participants (2/9) thought the number of messages was enough, while the other participants (3/9) either could not answer the question because they did not receive the message on their smartphone, or they did not give a clear answer. It is important to note that the number of messages a participant received depends on the number of times the participants docked the activity monitor.

Following this questions we asked participants what they considered to be a good number of messages a day. Five participants (5/9) could not answer this question, again because they did not receive the message on their smartphone, or because they could not put a number on it. One participant preferred not to receive any messages, while one participant would like to receive two or three messages a day. Another participant would like to receive a maximum of two messages a day, while the last participant would like to adjust the number of message a day herself.

Timing We asked participants what they thought of the timing of the messages. One participant thought the timing was good, while two participants (2/9) could not answer this question because they did not properly receive any messages on their smartphone. One participant thought the timing was ok, while the remaining five participants (5/9) thought the timing was bad.

We asked participants what they did not like about the timing of the messages, they answered the following:

What did you not like about the timing of the messages?
"I got a message at 21:00 which said 'You are well on your way to reach your target today', but at that point I would not do anything anymore."
"The messages are not 'smart' enough when it comes to timing."
"I got a message in the morning saying 'Well done, you are on your way to reach your target', which was fine. But then I got a message at 11:00 saying I should be more active. But I was at work, so that was not possible!"
"When I got back from a long walk I docked my activity monitor and got the message that I should be more active to reach my goal. But I was just back from a long walk!"
"Docking requests came at the wrong time, for example during diner." (3/9)
"I got all my messages after 18:00 which is too late."

After asking what participants did not like about the timing of the messages, we also asked them when they would like to receive messages, and when they did not like to receive any message. Results of these questions can be found in table 6.5.

Table 6.5: When participant would or would not like to receive messages

When would you like to receive messages?	**When would you NOT like to receive messages?**
8:00	after 20:00 (2/9)
16:00	after 22:00
22:00	while at work (3/9)
10:00, 14:00 and 17:00	during dinnertime
At lunchtime/during lunch break (5/9)	at night (2/9)
During the times of day when you take a break	I would like to have a button 'no messages today'
After work	
I don't want to get any message)	
By day	
In the afternoon	
In the evening	
More adjusted to the rhythm of the user (5/9)	

Content Two participants (2/9) did not receive any message on their smartphone; therefore they did not answer questions about content except when we asked them what they would like to receive as content. We asked participants what their general thoughts were about the content of the messages. Participants answered the following:

What is your general thought about the content of the messages?
"Kind of standard."
"Not very useful, when I look at the percentage I can see for myself how far along I am and if I have to do more."
"Bad, I thought the advice given would be more specific."
"Not really useful so far. I think I only got three different messages. 1 be more active, 2 well done, 3 dock your activity monitor."
"Bad. It gives little information. Even after just exercising I directly get a message to be more active."
"Message like 'try to be more active' are not very concrete."
"The messages do not give me enough information, which causes me to look more at the website."
"I think the messages are meaningless, and kind of administrative."
"I thought there were not enough different message, I think I got three different messages."

Secondly, we asked participants which messages they did or did not like. These results can be found in table 6.6. We also asked all participants (except the ones that did not receive messages on their smartphone) if they thought the content of the messages was correct. Five participants (5/7) answered 'Yes', one participant answered 'No', and one participant answered: 'not when I get a message at 21:00 that I am on my way to reach my goal, while I will not do anything anymore'.

Table 6.6: Which message did or did you not like?

Which messages did you like?	**Which messages did you not like?**
N/A (2/9)	N/A (2/9)
When you were active enough (2/9)	Getting a docking request when you already reached you target for that day
I did not like any message (5/9)	When the message did not take my situation into account
	Getting a docking request (2/9)
	You are doing a good job to reach your goal
	Nothing (2/9)
	Do something active in the next hours, when you just got back from a long walk
	"I did not like the wording, 'keep op the good work'."

Of course we also asked participants what they would like to have as content of the messages and how we could improve them.

What would you like to receive as content of a message, or how can the message be improved?
"I would like it if the messages were more specific, that they tell me how long I should go for a walk, and how many calories I burn doing that. Also I would like it if it gave me an alternative. For example, that if you do not feel like walking, the alternative is cycling." (2/9)
"When the system gives more specific messages, the content should be tailored to that specific user. The system should know which kind of activities the user likes, and take this into account when generating a message."
"I would like to receive a message when I reach 50% of my goal."
"I would like to receive tips for activity I can do while being at work in my office."
Messages should be concrete and specific: try to go for a walk of x minutes/km. The activity should be adjusted to the users' preferences. (5/9)
"It is hard to give an example of a message, because the data is not always synchronized."
"Messages like: 'you did not dock in the last two hours; last time you docked you did not reach your target. If you were active in these hours, please dock your activity monitor, otherwise go do something."'
"There should be more different kinds of messages."
"Messages should be more focused on the specific user."
"Messages and wording should be more divers."
"I would like to be able to incorporate exercise I got from my physiotherapist."
"Messages should be based on the normal activity pattern of the user. For example on a normal Monday the pattern is this, and now the user is less active than usual, then a message should be send with for example 'you are less active than normally try to be more active by doing xxx"'
"You should be able to make a list of activities that you think you can do on a day, so the system has a list to choose from when you are not active enough."

The system should be motivating; therefore we asked participants whether they thought the messages were motivating. None of the participants experienced the messages as being motivating. Reasons that were given include: the intonation of the speech, the content of the messages if mostly informing, messages are not specific or concrete enough, messages are annoying because they are not synchronized with the activity monitor, and messages are too late to be more active.

Subsequently, we asked how we could make the messages more motivating. Messages that were proposed are:

How can we make the messages more motivating, try to give an example?
"You have now burned x of x calories for today."
"You are at 20% of your target activity for today, while it is now 13:00. You should go do some physical activity. (do not send this at this time, when you know the users should be working now)"
"More specific message: 'Go walk/cycle/etc. for x minutes/km."
"When you walk/cycle/etc. for x minutes/km, you will burn x calories/ you will get x% closer to your target."
"When you have been more active compared to what you normally have done at that moment of the day, send a 'congratulations' message."
"When you have been less active compared to what you normally have done at that moment of the day, send a 'shame on you' message."
Something with group processes, but this is the only part of the interview which is unintelligible due to background noise in the recording.

There is no difference in the answers given by participants who used different versions.

6.3.5 Technology acceptance

In table 6.7 and table 6.8 all other measured constructs of technology acceptance are shown.

All constructs in table 6.7 are reliable (Cronbach's alpha > 0.6). The average scores lie under a neutral four for performance expectancy (PE) and usefulness (UN), while facilitating conditions (FC) scores neutral, and perceived ease of use (PEoU) just under neutral (between three and four). When we compare the scores of both versions we find no significant difference for all of these constructs (individual items or whole constructs). However, there are two items that have a notable larger difference between the two versions, being: 'Using the system I am more active.' of the PE, and 'Learning to operate the system would be easy for me.' of the PEoU.

We will now look at table 6.8. Only the construct of attitude (Att) is reliable (Cronbach's alpha > 0.6), while the constructs of trust (T), adaptability (Ada), and anxiety (Anx) are not reliable (Cronbach's alpha < 0.6). When we compare the scores of both versions we find no significant difference for all of these constructs (individual items or whole constructs if the construct is reliable). Looking at intention to use (IU) we see that although average scores lie well below average, there is one notable score of five in the ECA group, which is the only score above average. Scores for trust lie around average. There is one remarkable score, in which one participant scores 'I would trust the advice the system gives me.' high, indicating he would trust the advice, but scores 'I would follow the advice the system gives me.' low, indicating he would not follow the advice. Adaptability scores just below a average three, there are no notable scores. In anxiety a lower score is better due to the formulation of the statements. All items score well below average. There is one score that is remarkable for the item 'If I should use the system, I would be afraid to make mistakes of break something.' which scores a five. The construct attitude scores under a neutral four. There is one notable difference between the scoring of the item 'Working with the system is fun.', which is scored a full point high by the participants who used the text version of the system.

Table 6.7: Constructs: performance expectancy (PE), usefulness (UN), facilitating conditions (FC), perceived ease of use (PEoU).
Reliability is measured in the form of Cronbach's Alpha.
All items are measured on a 7-point Likert scale (1=strongly agree, 7=strongly disagree).

Construct	Statement	ECA group				Text group					Mean ECA group	Mean Text group	Reliability
PE	I find the system useful to be more active.	3	2	4	1	3	3	1	3	4	2.5	2.	0.68
	Using the system enables me to accomplish my goal to be more active.	3	5	4	1	3	1	1	2	2	3.3	1.8	
	Using the system I am more active.	3	5	2	1	3	3	1	5	4	2.8	3.2	
UN	Using the system would make it easier for me to be more active.	3	2	2	1	1	2	1	2	1	2	1.4	0.73
	I think the system useful for me.	3	5	4	1	3	4	1	4	3	3.3	3	
	I would find it useful to have the system.	2	6	4	1	4	3	1	2	4	3.3	2.8	
FC	I have the resources necessary to use the system.	5	6	3	5	7	4	1	3	4	4.8	3.8	0.62
	I have the knowledge necessary to use the system.	4	6	4	1	7	4	4	3	4	3.8	4.4	
PEoU	Learning to operate the system would be easy for me.	4	7	4	1	7	6	7	3	5	4	5.6	0.78
	I would find it easy to get the system to do what I want it to do.	4	4	4	1	2	4	4	3	2	3.3	3	
	My interaction with the system would be clear and understandable.	4	5	4	1	5	4	1	3	2	3.5	3	
	I would find the system to be flexible to interact with.	3	3	4	3	4	3	1	2	3	3.3	2.6	
	I would find the system easy to use.	5	6	4	1	5	4	4	2	4	4	3.8	

Table 6.8: Constructs: intention to use (IU), trust (T), adaptability (Ada), anxiety (Anx), attitude (Att).
Reliability is measured in the form of Cronbach's Alpha.
Items are measured on a 7-point Likert scale (1=strongly agree, 7=strongly disagree), apart from the items trust and adaptability, which are measured on a 5-point Likert scale (1=strongly agree, 5=strongly disagree).

Construct	Statement	ECA group				Text group					Mean ECA group	Mean Text group	Reliability
IU	Would you like to continue to use the system.	2	5	4	1	3	3	1	2	2	3	2.2	-
T	I would trust the advice the system gives me.	3	2	3	4	3	2	4	2	3	3	2.8	-3.6
	I would follow the advice the system gives me.	2	4	3	1	2	2	2	3	3	2.5	2.4	
Ada	I think the system will help me when I consider it to be necessary.	3	2	2	2	2	1	3	2	3	2.3	2.2	0.33
	I think the system will only do what I need at that particular moment.	3	2	3	2	3	1	3	2	3	2.5	2.4	
	I think the system can be adaptive to what I need.	2	2	4	3	2	3	3	4	3	2.8	3	
Anx	I am scared to use the system.	2	1	2	1	1	1	1	2	2	1.5	1.4	
	I find the system intimidating.	3	1	3	1	2	2	1	2	3	2	2	0.53
	If I should use the system, I would be afraid to make mistakes of break something.	2	1	1	1	2	2	1	5	1	1.3	2.2	
Att	I think it's a good idea to use the system.	3	5	4	1	2	2	3	6	4	3.3	3.4	0.91
	The system would make my life more interesting.	2	5	4	1	2	1	1	4	4	3	2.4	
	Working with the system is fun.	3	3	4	1	3	4	3	4	5	2.8	3.8	
	I like working with the system.	2	5	3	1	2	3	2	4	4	2.8	3	

6.3.6 Coaching

We used the source credibility scale of McCroskey (see section 4.6 and appendix A.6) of which we used the 12-items version (see appendix A.6.1. For the results of this questionnaire we refer to table 6.9

Table 6.9: Source credibility, 7-point scale

Construct	Statement	ECA group				Text group					Mean ECA	Mean Text	Reliability
Authoritativeness	unreliable–reliable	4	4	5	4	4	4	2	4	6	4.3	4	0.88
	uninformed–informed	4	3	4	2	4	4	3	2	6	3.3	3.8	
	unqualified–qualified	4	5	5	2	4	4	2	4	6	4	4	
	unintelligent–intelligent	3	3	3	4	4	4	2	5	6	3.3	4.2	
	worthless–valuable	3	4	5	1	4	4	2	4	6	3.3	4	
	inexpert–expert	3	3	3	4	4	4	4	2	6	3.3	4	
Character	dishonest–honest	5	2	4	4	4	4	6	4	6	3.8	4.8	0.65
	unfriendly–friendly	4	6	4	4	4	4	4	4	5	4.5	4.2	
	unpleasant–pleasant	4	3	5	2	4	4	1	4	6	3.5	3.8	
	unselfish–selfish	5	6	4	4	4	4	4	4	6	4.8	4.4	
	awful–nice	3	5	4	4	4	4	2	4	5	4	3.8	
	sinful–virtuous	4	4	5	4	4	4	4	4	6	4.3	4.4	

Reliability of each construct of the source credibility is good (Cronbach's alpha $>$ 0.6). We see that the mean scores for each item lie around a neutral four. When we look as the mean score of each construct per version we see a small difference for autoritativeness (3.6 ECA vs. 4.0 Text). For character these means are equal (4.0 ECA vs. 4.0 Text). We find no significant difference between the two groups (for the whole construct or for individual items).

For the results about the quality of coaching we refer to table 6.10.

Reliability of each construct of the quality of coaching questionnaire is good (Cronbach's alpha > 0.6) for the first construct (mental preparation) and the third construct (other). Reliability of the second construct (goal settings) is not good. When we look at all scores given it is notable that almost all scores lie below a neutral three (most scores given are one or two). When we look as the mean score of each construct per version we see a small difference for mental preparation (1.8 ECA vs. 1.3 Text) and other (1.8 ECA vs. 1.4 Text).

We only found one significant difference in the quality of coaching's questionnaire. This difference is found in the statement: "My coach helps me identify strategies to achieve my goals" (see table 6.10). ECA group: M = 2.8, SD = 0.96, Text group: M = 1.4, SD = 0.55, the difference was significant $t(7) = 2.68$, $p = 0.032$. Looking at the individual scoring given to the statement, we see a prominent difference between the scoring for all participants who had a fully working system and the two participants who did not receive their messages on the smartphone correctly. All participants gave a score of one or two to this statement while the participants without a fully working system scored this statement with a three and four. At two other statements we found an almost significant difference between the two groups. The statement "My coach helps

Table 6.10: Coaching Scores (5-point scale (1=strongly agree, 5=strongly disagree)), first construct: mental preparation, second construct: goal settings, third construct: other. Reliability is measured in the form of Cronbach's Alpha

Statement	ECA group				Text group					Mean ECA	Mean Text	Reliability
My coach gives me advice on how to be mentally tough.	2	1	3	1	2	1	1	2	1	1.8	1.4	
My coach gives me advice on how to stay confident about my abilities.	2	2	2	1	1	1	1	2	1	1.8	1.2	0.90
My coach gives me advice on how to stay positive about myself.	3	1	3	1	1	1	1	2	1	2	1.2	
My coach gives me advice on how to stay focused (on my activity goals).	2	1	2	1	2	1	1	2	1	1.5	1.4	
My coach helps me identify strategies to achieve my goals.	2	2	3	4	2	1	1	2	1	2.8	1.4	
My coach monitors my progress toward my goals.	3	4	2	1	1	1	1	2	5	2.5	2	
My coach helps me set-short term goals.	2	1	3	1	2	1	1	2	1	1.8	1.4	0.37
My coach helps me identify target dates for attaining my goals.	2	2	3	4	2	2	1	2	1	2.8	1.6	
My coach provides support to attain my goals.	3	2	2	1	1	1	1	2	1	2	1.2	
My coach helps me to recognize and celebrate my activity achievements.	2	2	2	1	1	1	3	2	1	1.8	1.6	
My coach helps me to be motivated and inspired by others.	2	1	2	1	1	1	1	2	1	1.5	1.2	
My coach helps me to discover which things help me to become more active.	2	2	2	1	1	1	1	2	1	1.8	1.2	0.85
My coach had the right knowledge and abilities to give good coaching.	2	2	3	1	2	1	1	2	1	2	1.4	
My coach gives coaching of good quality.	2	2	3	1	2	2	1	3	1	2	1.8	

me identify target dates for attaining my goals", ECA group: M = 2.8, SD = 0.96, Text group: M = 1.6, SD = 0.55, the difference was almost significant $t(7) = 2.28$, $p = 0.056$. When we look at the individual scoring of this statement we again see that the two participants without a fully working system scored this statement different than those with a fully working system. Finally, we find an almost significant difference in a statement asked as expectation: "The system would make my life more interesting", ECA group: M = 3.0, SD = 1.41, Text group: M = 1.4, SD = 0.55, the difference was almost significant $t(7) = 2.35$, $p = 0.051$.

6.3.7 Privacy

Participants were asked whether they had any objection to the information that is gathered by the system (activity data). All participants (9/9) indicated that they did not have any objection to the gathering of the activity data.

We asked participant if they would have any objection if location (GPS) was added to the gathered information (to be used to time messages better and to get a more appropriate content). Three participants (3/9) answered they still did not have any objection. One participant only did not object if the data was used for himself to see the distance he had walked for example, and if he could turn location data off. Three participants (3/9) would have objections if location was used. Two other participants (2/9) would not have any objection if the data was not stored.

A participant proposed a connection with his agenda to be used for the same reasons location data would be used. Three participants (3/9) thought this was a good idea. One participant had no objection if the agenda was separated, and no connection existed

between the system and his Google-calendar. Two participants (2/9) had no objection when they system only could see which hours were occupied or if the agenda was separate. When the user has control over what the system knows from the agenda, two participants (2/9) would not have an objection. In general, users do not want to have the feeling 'Big-brother is watching you'; this was explicitly said by two participants (2/9).

There is no difference in the answers given by participants who used different versions.

6.3.8 User experience

User experience was measured with the Attrakdiff questionnaire of Hassenzahl (for more information about Attrakdiff see section 4.3 or appendix A.3). The results for the ECA group can be found in figure 6.4, while the results of the text group can be found in figure 6.5. Finally, we combined both groups, these results can be found in figure 6.6.

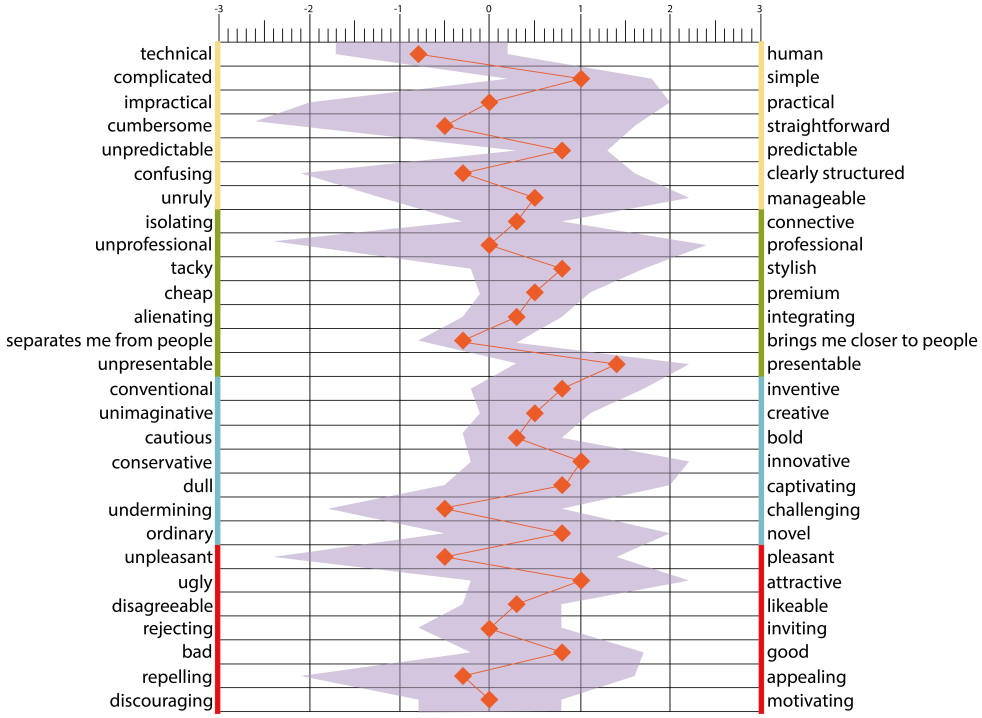


Figure 6.4: Attrakdiff results ECA group
pragmatic quality (yellow) Cronbach's $\alpha = 0.86$, hedonic quality - identity (green) Cronbach's $\alpha = 0.71$, hedonic quality - stimulation (blue) Cronbach's $\alpha = 0.51$, and attractiveness (red) Cronbach's $\alpha = 0.87$
orange line = mean value, greyish area = standard deviation

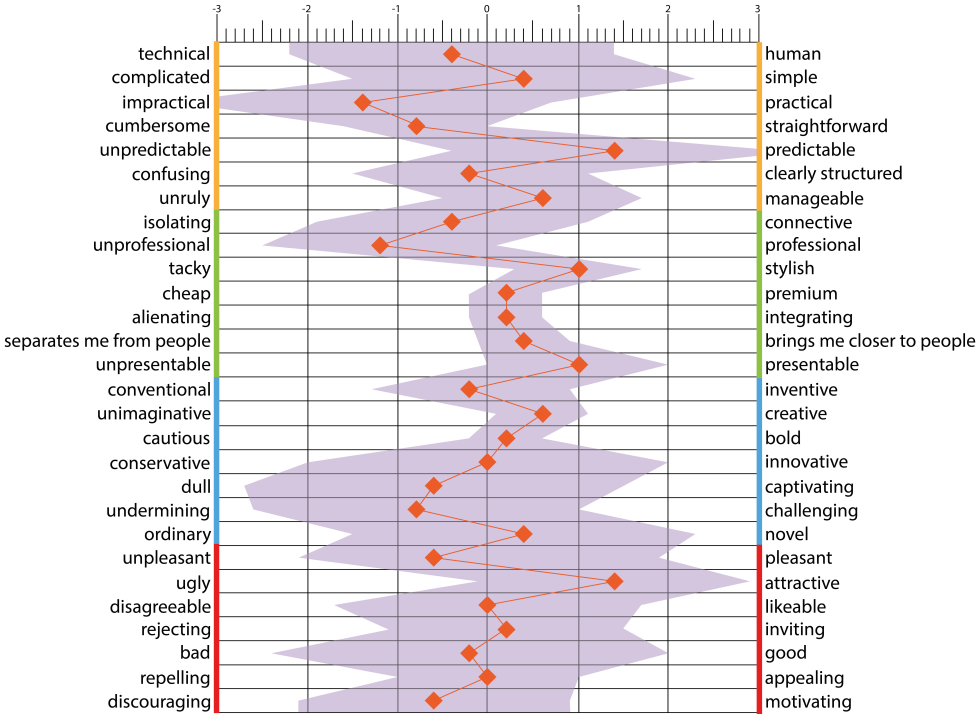


Figure 6.5: Attrakdiff results Text group
pragmatic quality (yellow) Cronbach's $\alpha = 0.81$, hedonic quality - identity (green) Cronbach's $\alpha = 0.55$, hedonic quality - stimulation (blue) Cronbach's $\alpha = 0.87$, and attractiveness (red) Cronbach's $\alpha = 0.76$
orange line = mean value, greyish area = standard deviation

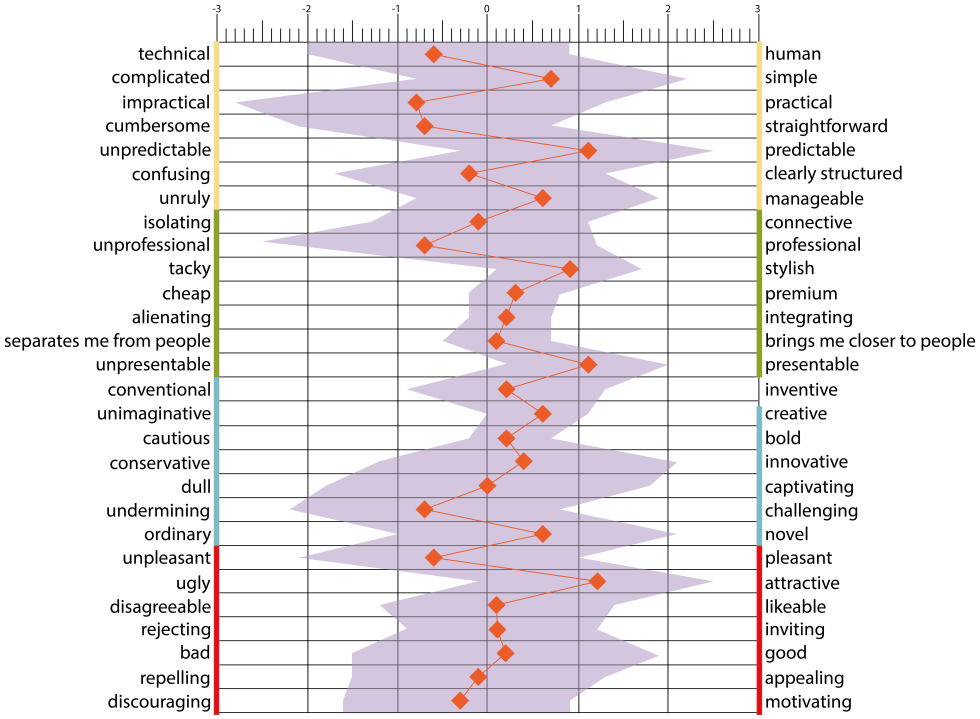


Figure 6.6: Attrakdiff results all participants (ECA and text group combined)
pragmatic quality (yellow) Cronbach's $\alpha = 0.82$, hedonic quality - identity (green) Cronbach's $\alpha = 0.62$, hedonic quality - stimulation (blue) Cronbach's $\alpha = 0.83$, and attractiveness (red) Cronbach's $\alpha = 0.80$
orange line = mean value, greyish area = standard deviation

We will first look the results of the ECA version, after which we will discuss the text version. Finally, we will look at the combined scores.

When we look at figure 6.4 we see that in general, word pairs score average (average being between -1 and 1). The only word pair that scores above average is unpresentable–presentable. The reliability of each dimension is good (Cronbach's alpha > 0.6), apart from the hedonic quality - stimulation. When we remove undermining–novel from this dimension it becomes reliable with Cronbach's alpha $= 0.86$.

The average values of each dimension are shown in figure 6.7 (orange). We already saw that all dimensions, apart from the hedonic quality - stimulation were reliable when taking all word pairs into account; therefore we should be careful interpreting the average value hedonic quality - stimulation. All averages lie around average.

When we look at figure 6.5 we see that in general, word pairs score average (average being between -1 and 1). There are two word pairs which score below average (impractical–practical and unprofessional–professional), and two word pairs which score above average (unpredicatable–predictable and ugly–attractive). The reliability of each dimension is good (Cronbach's alpha > 0.6), apart from the hedonic quality - identity. When we remove isolating–connective from this dimension it becomes reliable with Cronbach's alpha $= 0.70$.

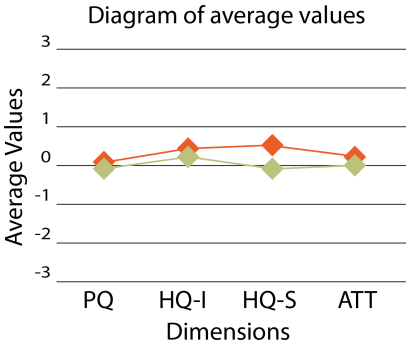


Figure 6.7: Attrakdiff results averages for each construct (ECA version = orange, Text version = green)

The average values of each dimension are shown in figure 6.7 (green). We already saw that all dimension, apart form the hedonic quality - identity were reliable when taking all word pairs into account; therefore we should be careful interpreting the average value hedonic quality - identity. All averages lie around average.

When we look at figure 6.6 we see that in general, word pairs score average (average being between -1 and 1). There are three word pairs which score above average (unpredicatable–predictable, unpresentable–presentable and ugly–attractive). The reliability of each dimension is good (Cronbach's alpha > 0.6).

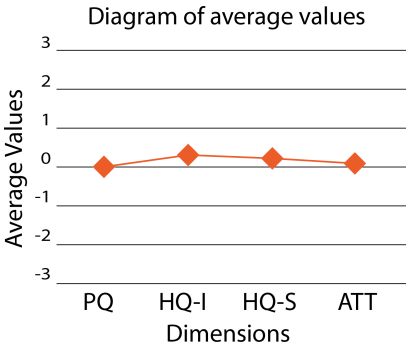


Figure 6.8: Attrakdiff results averages for each construct (both groups combined)

The average values of each dimension are shown in figure 6.8. All averages lie around average. We already saw that all dimensions were reliable when taking all word pairs into account; therefore we can treat the whole dimension for what it is supposed to measure.

We found no significant difference between the two versions, not when we compare individual items, or when we compare reliable constructs.

6.3.9 Usability

We asked our participants to fill out the System Usability Scale (for more information about the system usability scale see section 4.1 or appendix A.1). The results can be found in table 6.11. The System Usability Scale is reliable with Cronbach's alpha $=$ 0.86, and therefore reliable. System usability scores can range from 0 till 100. When

we look at the scores given to the system we see that the lowest score is 20 while the highest score is 77.5. The mean score is 51. There is a large variation in scores.

There is no significant difference in the scores given by participants who used different versions (in the whole usability construct or individual items).

Table 6.11: Usability Scores (5-point scale (1=strongly agree, 5=strongly disagree))

Statement	**ECA group**				**Text group**					**Mean score ECA**	**Mean score Text**
I think that I would like to use this system frequently	2	4	4	1	2	2	1	3	4	2.8	2.4
I found the system unnecessarily complex	2	1	2	5	2	2	1	4	2	2.5	2.2
I thought the system was easy to use	4	4	4	1	3	4	3	2	2	3.3	2.8
I think that I would need the support of a technical person to be able to use this system	2	1	2	1	1	1	1	4	2	1.5	1.8
I found the various functions in this system were well integrated	3	3	3	1	1	2	1	2	3	2.5	1.8
I thought there was too much inconsistency in this system	3	2	3	5	4	4	3	5	2	3.3	3.6
I would imagine that most people would learn to use this system very quickly	4	4	4	1	5	4	3	3	4	3.3	3.8
I found the system very cumbersome to use	2	2	2	5	4	4	3	4	2	2.8	3.4
I felt very confident using the system	2	3	3	1	3	1	1	2	4	2.3	2.2
I needed to learn a lot of things before I could get going with this system	2	1	2	1	1	1	1	3	2	1.5	1.6
System Usability Score	60	77.5	67.5	20	55	52.5	50	30	67.5	56.3	51

6.3.10 Other

Other remarks that were made that did not answer the asked question, but had a different subject included the following:

"It would be better if you did not have to dock the activity monitor and it would synchronize automatically."

"Messages should be tailored to your specific activity pattern."

"Messages should be more personal."

"The website was too slow."

"The system should be 'smarter'."

"The feedback given by the system should be improved."

"It should be possible to adjust the number of message yourself."

"Messages should be more specific."

"The messages are too much alike."

"I did not trust the system."

"I would like to have a heart rate monitor added to the system."

During the interview we explained that there were two versions of the system (ECA version and Text version). We asked participants which they preferred, although they

had not experienced both (only participants who had the text version or not a fully working ECA version have answered this question). Three participants (3/9) preferred the text version, while one participant had no preference. Reasons and reactions included, 'I would find the ECA too childish', 'I do think an ECA would be more personal', 'I think an ECA would be more motivating', 'An ECA would remind me of my navigating system or the paperclip of Word', 'Text is less personal', 'When the ECA shows emotion it will be more personal', 'I prefer to be able to see the message without any sound'.

6.4 Discussion SmarcoS-office worker

In the section we will discuss the results of the SmarcoS-office worker evaluation. We will take a closer look at the results and point out some interesting findings. We will look at the results, and discuss how they may be explained. Finally, we will discuss how these results give us more insight in behaviour change support systems like the SmarcoS system.

Expectations When we asked participants about their expectations (in their own wording), we expected to see slightly more positive expectations. Only four participants (4/9) have expectations that correspond with the goals of the system. All other participants either do not have expectations or have more negative expectations. This shows that beforehand, without knowing the system, participants are somewhat hesitating towards the system. This may be partly due to the fact that our participants did not represent our target group the best they could. All were office workers, but it is debatable to which extend they were intrinsic motivated to be more active. Secondly, it also is debatable whether the participants that had the intrinsic motivation where in the right stage of behaviour change (see section 3.1.1, Transtheoretical model of Behaviour Change) for the system to work. We will look more into this in the next part 'Using the system'. A more simple explanation may be that they knew it was a system that was just developed, and therefore were hesitating, or because some had experience with behaviour change support systems.

The ratings that were given to the statements were not that high, when you keep in mind that most statements were measure on a seven-point scale. Most ratings lie below a neutral score of four. When we compare the scores given to the same statements after using the system for a couple of days we see that the scoring of the statements afterwards is even lower. In the case of anxiety this is a good thing. However, in all other cases this means that while the expectations were already not that high, in reality the system even performed worse. This is not true for the statement "The system would make my life more interesting." for both groups, and the statement "Using the system enables me to accomplish my goal to be more active." for the ECA group. In the last case, this was not caused by the scoring of the two participants who did not receive the message on their smartphone properly (individual scoring afterwards can be found in table 6.7). In short, the system did not meet the already low expectation of participants. Reasons why the system did not meet expectations will become apparent in the remaining of the discussion.

Using the system Participants indicated that they would not use the system as it currently is. However, they see potential value in it as is illustrated by the fact that most participants think such a system could help people live a more healthy and balanced life, but only if the users want to change. Hereby they also provide a reason why they would not use the system, being that they do not really want to be more active. This is said by multiple participants during the interviews. Secondly, as already stated before, it is debatable to which extent participants had intrinsic motivation to be more active. A

behaviour change support system only works if the user is willing to change his behaviour and if he is in the fourth stage of behaviour change (see section 3.1.1, Transtheoretical model of Behaviour Change). The fourth stage of behaviour change is the action stage in which people actually make modifications to their life. Most participants were in one of the previous stages of behaviour change (precontemplation, contemplation or preparation) or in no stage if you would argue that when you are active enough and do not need to be more active you are not in a stage. Most of our participant already were quite active outside of their work, which is illustrated by the list of sport they report doing. Their motivation to be more active is therefore limited.

In the measured constructs of technology acceptance, we also asked a question on the intention to use. There was one participant who scored this item more positive than other participants. We see the same in his answer during the interview, which is more positive in comparison with other participants. It is also notable that this participant seemed to be more willing to be more active in comparison with other participant. This shows that the stage of behaviour change a user is in, is really important.

When we look at the changes participants proposed in order for them to want to use the system it is apparent that most suggestions concern functionality. This gives an indication that the functionality of the system is not meeting their basic standard. Changes to improve functionality will be discussed further on.

System information and functionality The time it takes to use the system is seen by the participants as fine. However, user friendliness can be improved.

Installing all applications and adjusting the settings is seen as not user-friendly by participants of the ECA-group. This is understandable since they had to install three applications and adjust multiple setting. For an evaluation this is not a problem, unless it leads to a not fully functioning system as is what happened with two participants. This can be prevented by better informing the participant how the system should work when it is functioning correctly. It was stated in each document and questionnaire that when participants had problems or questions they would be welcome to ask them. Although participants noticed something was wrong, none of the participants who had problems with the application did report this before the end interview. By better informing them how a fully working system looks like and how it functions, it might help to let them ask for advice or help. Of course, this also means that in the 'final' system installing the application and adjusting setting should be simple and clear. Secondly, the loading speed of the DirectLife website was also not considered user-friendly. The loading speed was too slow even for evaluation purposes, because it has influenced the use of the system. This was illustrated by a participant who said that she probably did not use the website properly, due to it being too slow. Therefore, loading speed of the website should be increased. To which extend the loading speed has influence the use of the system is hard to say. We cannot compare use of the application with use of the website. Because we have no data of the use of the website, and not all functionality in which participants were interested could be seen on both. An example of this is the activity pattern during the day, which only could be seen on the website.

Errors reported by participants were mainly functional errors. This will have influenced the use of the system, and the experience with the system. These errors also show that content of messages can be seen as an error, while it does follow the rules defined. Example of this were: 'I got a docking request when I already reached my target for that day.' and 'I got a 'You are doing well' message at 8:00 while I only got dressed and drove my car to work.'. Therefore, some rules should be revised.

The most important functionality that should be included according to participants is, omitting the docking of the activity monitor and synchronizing it automatically (five participants (5/9) gave this as part of the answer to this question). Implementing this

will drastically change the system, because activity data will be readily available. This will also influence content and timing of the messages which we will discuss next.

Activity messages When we look at all results concerning activity messages we see a pattern in preferences of the participants. Participants would prefer it when the number of messages and the timing of the messages could be adjusted to the preferences of the participant. Mainly because what suits one participant is completely wrong for the other participant. This is illustrated by the different suggestions participants give for the timing of the messages (table 6.5). Lunchtime is mentioned by many participants as a good time to receive a message, but the chance that they all are having lunch at the same time is small. Therefore, this timing should be adaptable for each participant.

The content of the message was mostly seen as standard, administrative, not motivating, not divers enough and there were not enough different types of messages. When we look at the comments given to the question 'What is your general thought about the content of the messages?' we find no positive comment. And when we specifically asked them which messages they liked, only two participants answered: getting a message when you were active enough. But this was said more as result of being active enough than they liked the message. In short, participants think the content is boring and not motivating. This is a problem since the main goal of the system is to motivate office workers to be more active and life a more healthy and balance live. When messages are seen as boring and not motivating, then how can your system be motivating?

Participants agree with each other on how to improve the content of the messages. Messages need to be more specific, more concrete, and more personal. When users are not active enough the messages should not only tell them that they should be more active, it should also say what the user can do to be more active. It would be even better to specifically suggest an activity, combined with the duration of that activity and what the result afterwards will be. For example: 'You are not activity enough today. To be more active you can go for a walk. When you take a 30 minute walk, you will burn 222 calories. This brings you 30% closer to your target. After your walk you then will be at 60% of your target!'. The content of the message should be more tailored to the preferences of the user, the system should take into account which activities the participant likes to do, and maybe even if they are feasible in the normal routine of the user (as was suggested by one participant). Improving the content of the messages this way will also increase how motivating the messages are according to the participants.

Adjusting timing and the number of message with the system as it is right now might be difficult, because feedback can only be given accurately and relevant if the user has just docked their activity monitor. Part of the reason why messages are not seen as being relevant has to do with this problem, as also was noticed by multiple participants.

Technology acceptance The reliability of some of the constructs measuring technology acceptance (performance expectancy, usefulness, facilitating conditions, perceived ease of use, intention to use, trust, adaptability, anxiety, and attitude) is not that good. For performance expectancy, usefulness, facilitating conditions, perceived ease of use and attitude Cronbach's alphas are high enough (0.68, 0.73, 0.62, 0.78, 0.91) when we keep in mind that we had small group of participants. Reliability of trust, adaptability and anxiety are too low therefore we will treat each item of these construct individually. We found no significant differences in acceptance between the two version of the SmarcoS-office worker system (not for whole constructs, if applicable, or for separate items).

Most constructs scored below average (average being four if measured on a seven-point scale, average being three if measured on a five-point scale). Exceptions to this are the construct facilitating conditions, the first and last item of perceived ease of use

('Learning to operate the system would be easy for me.' & 'I would find the system easy to use.'). While the first item of trust and last item of adaptability score just barely on average. We will shortly discuss all constructs.

Performance expectancy scores below average/neutral indicating that performance of the system was not as good as expected. This is in line with all other results. Usefulness scores well below average, indicating that participants did not find the system useful. We see the same in other results. Facilitating conditions scores on average, which tells us that participants had the resources and knowledge that were necessary to use the system. Perceived ease of use is somewhat divided. As already state the first item 'Learning to operate the system would be easy for me.' scores above average, therefore learning to use the system is not too difficult for participants. The last item 'I would find the system easy to use.' tells us that the system is at not too hard to use. However, all other items score below average this tells us that further improvement is needed. Intention to use is already discussed in 'Using the system'. Trust items scores just below average indicating that participants do not wholly trust advice the system gives them and will probably not follow the advice. We see the same in the questionnaires concerning coaching. Adaptability of the system needs improvement based on these scores. We should treat these items individually based on reliability score, however all items need improvement. Anxiety items scores well below average. This is good because you don't want users to be anxious using the system. Secondly, it is even more positive because scoring of anxiety actually lowered in comparison with expectations beforehand. Finally, attitude also scores below average, indicating the attitude towards the system is not that good. We see the same in answers of the interviews.

Coaching We asked participants how they perceived the source credibility, and what they thought of the quality of coaching they received from the system. For source credibility scoring lies around a neutral four, although not too bad, this needs improvement. The scoring of the quality of coaching scoring lies below a neutral three, indicating that the quality of the coaching participants got was not good. This is mostly due to all previously mentioned reasons, with main reasons being the bad timing and content of the messages, since they form the majority of the coaching. We would like to argue that when the content and timing of the messages is improved, both source credibility and the quality of coaching will improve.

Privacy Privacy is always an important issue. How much privacy are users willing to give up, in exchange of a better working system? The data gathering of the system now was only limited to the activity data. Although, as argued before, we think this data was not optimally used, adding other data might improve the system. None of our participants indicated having a problem with gathering of the activity data, mainly because it is not seen as too personal. When we suggested the adding of location or access to some kind of agenda, to help improve messages, we found a border. Three participants (3/9) did not have objection to adding location data, while others were more hesitant wanting to have either some control over the data, or demanding it would not be stored and only used. The same pattern was seen with the suggestion of using the content of an agenda to improve messages. By using data like location or data from an agenda participants associated this with a feeling of 'Big-brother is watching you'. This data was seen as more personal than the activity data, and explains why participants were more hesitant. These results are different from what we found in the SmarcoS-diabetic evaluation. This will be discussed in section 7.1.

User experience We see that most word pairs of the Attrakdiff questionnaire score around zero. We already stated there was no significant difference between the two

versions of the system, therefore we will only discuss results of the combined systems (figure 6.6). The averages of each construct (figure 6.6) lie in the average region (the average region lies between minus one and one). All scores are not extreme enough to give us a clear picture of how participants perceived the different dimensions. Therefore, we should treat these results more as a trend, and be aware that all dimensions could benefit from improvements. We will now discuss all dimensions, beginning with pragmatic quality.

Looking at the pragmatic quality (yellow beam in figure 6.6), participants find the system kind of technical, impractical and cumbersome, but also quite simple and predicable (most notable scorings). This more or less corresponds with the other questionnaires and the interview. Usability of the system should be improved (as mentioned before) and users are not really successful in achieving their goals using the system (of course with the remark that it already is debatable to which extent participants were motivated to become more active).

Participants find the system presentable, when we look at hedonic quality - identity (green beam in figure 6.6). This positive because, users are more lenient towards a presentable system. When we would improve hedonic quality - identity for the system, users can identify themselves better with the system. We think this is important since the system should be used to attain your *personal* goal of becoming more active.

When we look at the hedonic quality - stimulation (blue beam in figure 6.6), there is one word pair that clear scores different than all other word pairs (lower) in this construct being undermining–challenging (indicating participants find the system undermining). It is debatable to which extend this system should be challenging. Attaining the goal of becoming more active can be challenging, but using the system should not be challenging. Therefore, we would like to argue that although this score is low, it is not negative for the system.

Participants find the system attractive but also unpleasant, as is indicated by the scoring of the attractiveness dimension (red beam in figure 6.6). We also see that participants do not score the system as being motivating, which is something we would want. An attractive system will have greater persuasive power than an unattractive technology, and the mere appearance of a system is sufficient to change its social influence [31, 74]. As is also stated in section 3.1.3.

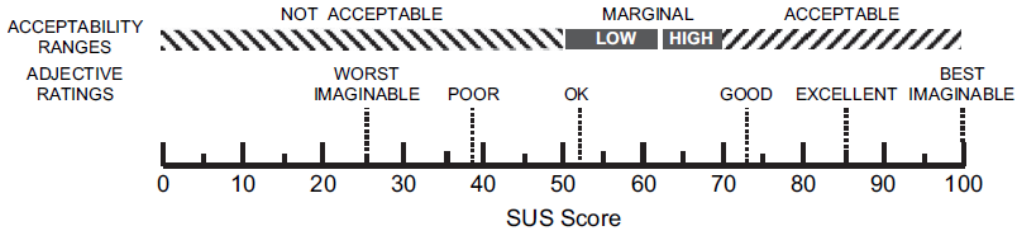


Figure 6.9: Adjective ratings, and acceptability of the overall SUS score

Usability We already discussed in section 4.1 what makes a good system usability score (see figure 6.9). We established that a system usability score of 70 or higher is good, while we are looking for a score in the high 70s or even higher. When we look at the scores given by participants we see that the system usability score varies between 20 and 77.5. There is just one participant who gave the system a good system usability score. All other participants scored the system well below the 'good' score. The average score of the SmarcoS-office worker-ECA group is 56.3 what would be marked as just

ok. While the average score of the SmarcoS-office worker-text group is even lower of 51, which would be just below ok. There are no scorings which really standout, although most participants do not seem to find the system very hard to use. We would like to argue that the whole system should be improved, focusing on the functionality, since this is most prominent in other parts of the evaluation.

ECA vs. Text Comparing the two version of the SmarcoS-office worker system we barely see any differences in answers giving by participants during the interview, open questions of the questionnaires and ratings given to the statements.

In all 64 measured statements (excluding the Attrakdiff and SUS questionnaires) there was only one statement that was rated significantly different between the two groups. This difference was caused by the two participants who did not receive the messages on their smartphone correctly. Omitting these participants we found no significant difference. This was also the cause of the almost significant difference in the statement "My coach helps me identify target dates for attaining my goals.". The final statement that is *almost* significant ("The system would make my life more interesting.") is measured as expectation, while there was *no* significant difference in the rating afterwards.

Comparing the two groups in answers of the Attrakdiff questionnaire we also found no significant difference. Neither for the results of the System Usability scale. This tells us that there was no measurable difference in any of the measurements we did.

Based on other studies like [11, 74] we would expect a difference to occur. However, when we look at the interviews we see that participants are focused on giving suggestions improving functionality of the system. Therefore, it is likely that the functioning of the system is of more influence now than the difference between the SmarcoS-office worker-ECA version and the SmarcoS-office worker-text version. To be able to measure if there is a difference between the two versions we would suggest first to improve the functioning of the whole system, before focussing on the difference.

7 Discussion

7.1 Comparison of all versions of the SmarcoS-system

In the section we will shortly discuss the differences found in all evaluations. There are some differences between all versions of the system, which make it hard to compare them. First of all, the target group of the versions differs (diabetics vs. office workers). This is important to note because some differences we found can be explained by this. Secondly, some functionality is not the same. Therefore, we should be really careful when making comparisons between the versions. The different versions of the system do no function the same, but because of the shared basis (see section 5.1.1) we can make a careful comparison.

When we look for example at privacy, which is unrelated to any target group or version of the system, we see a clear difference. Participants of the SmarcoS-diabetic evaluation are more willing to give up privacy than participants of the SmarcoS-office worker evaluation. Yes, we still see the same boundary occur, but participants of the SmarcoS-diabetic evaluation are more willing to give up some of their privacy. This difference cannot be caused by the difference in systems, since they used the same data (to be more precise SmarcoS-diabetic system uses even more data in the form of medication intake). Because we can rule out that this difference is caused by the differences between the versions, the only other possibility is that it is caused by the difference in participants.

The main difference between participants is that participants of the SmarcoS-diabetic evaluation are diabetics, while the participants of the SmarcoS-office worker evaluation are office workers. The main difference between these two groups is that diabetics have a disease and therefore can be considered patients. The office workers had no disease, and therefore could not be considered patients. This difference also makes it likely that the participants of the SmarcoS-diabetic evaluation are more conscious that they need to have a healthy lifestyle, and take their medication on time. Therefore, they probably have a bigger interest or at least are more conscious of their interest in such behaviour change support system. It is therefore likely that they are more willing to give up privacy, because they get more advantages of giving up privacy compared to the non-patient office workers.

Another important difference between the two participant groups is that the participants of the SmarcoS-diabetic evaluation probably will notice a more direct effect of not having a healthy lifestyle, and when they do not take their medication on time. This is probably less direct for the participants of the SmarcoS-office worker evaluation, because the benefits of being active enough are more of a long term effect for them. Participants of the SmarcoS-diabetic evaluation probably will notice it much faster if they do not take their medication on time and are not active enough. This is also a reason why it would be more apparent for participants of the SmarcoS-diabetic evaluation why they system can help them, and why improving the system is more important.

When we look for example at the results of the Attrakdiff questionnaire, we see that the SmarcoS-diabetic system scores better than the SmarcoS-office worker system at all word-pairs except at unpredictable–predictable and ugly–attractive at the positive site. On these pairs the SmarcoS-office worker system scores slightly better. On the negative side the word-pairs of separates me from people–brings me closer to people, cautious–bold and undermining–challenging scores slightly better of the SmarcoS-office worker system. In general, the difference in scoring is very noticeable, while the differences

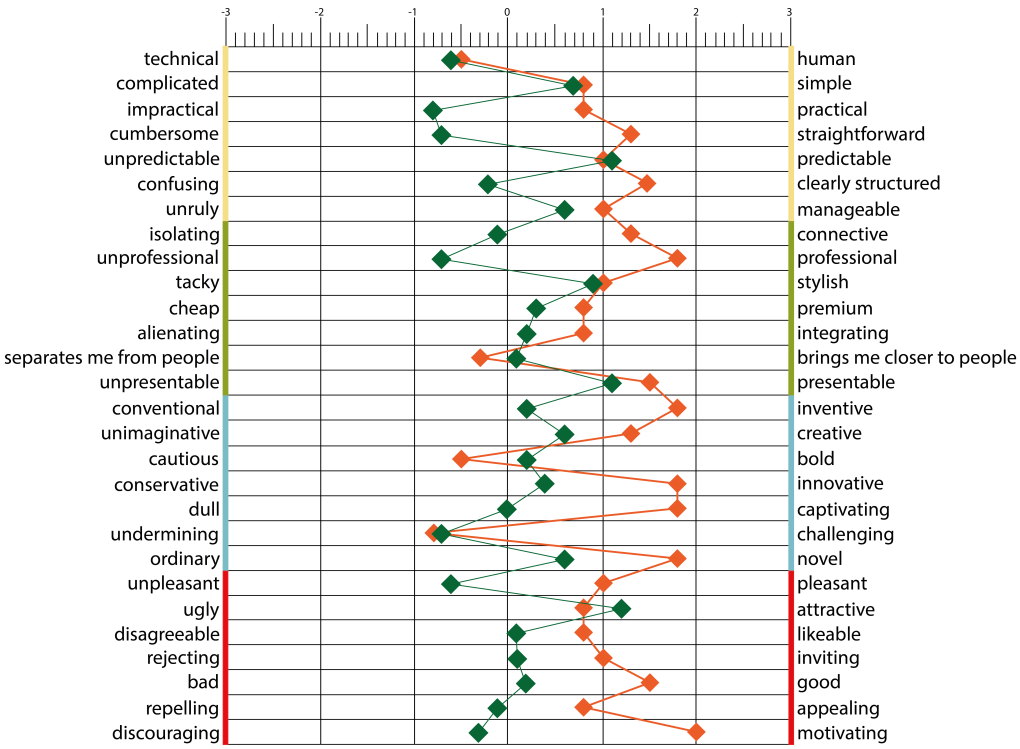


Figure 7.1: Attrakdiff results SmarcoS-diabetics (orange), SmarcoS-office worker (green)

between the systems are not that big. This can partially be explained by two factors. The first factor is that the target group is different. There is an important difference between 'patients' and 'non-patients'. We think 'patients' might be more willing to cope with lesser functionality, as long as a system helps them, without jeopardizing their health. While 'non-patients' want higher standards before even considering using a system. The need for a system is high for 'patients'. Secondly, and this is also related to the first factor, is that in order for a system to work users have to be in the correct stage of behaviour change (see section 3.1.1). Because the participants in the first evaluation were probably more aware of the need to change behaviours (in their case, medication intake and being more active) it is also more likely that they were at the right stage for the system to work. As noted by participants of the SmarcoS-office worker evaluation, the system will only help people live a more healthy and balanced life if people are willing to change. The will to change was greater in the SmarcoS-diabetics evaluation.

SmarcoS-diabetic participants had less experience with the activity messages and activity monitor. However, we see one of the most important comments of the SmarcoS-office worker evaluation also occur in the SmarcoS-diabetic evaluation: the activity monitor should synchronize automatically, omitting the need to dock. Secondly, was also said by a participant on the SmarcoS-diabetic evaluation: "Do not only say you do not exercise enough, but also give tips on what someone can do to be more active. Specific to the situation, if someone does not take his medication four days in a row, then something is wrong. Make sure you do something with that information, and give tips according to what is happening.". The specific-ness, the tips, and more focusing on the situation is exactly what is mostly said about the activity messages in the SmarcoS-office worker evaluation.

Finally, we can compare the results of the system usability scale (see section4.1, and appendix A.1). Of all comparisons we are now making this is the one that we should be

most careful and reserved about. The privacy issues were quite unrelated to the system. Attrakdiff measures user experience, and while the systems are not the same they are more or less comparable on this level without causing wrong implications. Usability is focused on the extent to which the system can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use. When we look at the difference in the scoring of the system usability scale we immediately see a large difference. The majority of the participants of the SmarcoS-diabetic evaluation score the system usability quite high (positive), as opposed to the SmarcoS-office worker evaluation in which only one participant scores the system 'good'. When we compare the willingness of participants to use the system after evaluation we see the same. How can we explain this? We would like to argue that this is due to several factors already discussed. First of all, there is a difference between 'patients' and 'non-patients'. Secondly, there probably is a difference between the stage of behaviour change between the two participant groups, likely caused by being a patient and the increased importance of living a healthy life. Finally, the difference can be explained by the difference between the systems, and the fact that the most important feature of the SmarcoS-office worker version (activity feedback), was of less importance in the SmarcoS-diabetic version of the system. The participants of SmarcoS-diabetic evaluation experienced less feedback on activity. Therefore, problems and missing functionality also were less prominent.

7.2 In relation to the theory

In this section we will look at our results in relation to the theory presented in section 3.1. In this section we discussed two psychological theories on which behaviour support systems build, and presented some important elements of a behaviour change support system. We will first look at our results in relation to the psychological theories presented, after which we will look at identified elements and make a comparison with the tested versions of the system. The focus hereby will lie on the part of the system that is concerned with improving the activity level of the user, since this is the shared feature for all versions of the system.

Transtheoretical model of Behaviour Change In the previous section we determined that most participants probably were not at the right stage (stage four: 'action') of behaviour change for the system to be most successful. As noted by participants of the SmarcoS-office worker evaluation, the system will only help people live a more healthy and balanced life if people are willing to change. Participants in the SmarcoS-diabetic evaluation were probably more aware of the need to change behaviours (in their case, medication intake and being more active) than the participants of the SmarcoS-office worker evaluation. Therefore, it is also more likely that participants of the SmarcoS-diabetic evaluation were at the right stage for the system to work. We have already discussed this in previous sections, and therefore will not repeat this.

Goal-Setting Theory The other psychological theory we look at is the goal-setting theory. It states that there is a relationship between the difficulty and specificity of a goal and the performance of the task. Specific, difficult goals consistently lead to higher performance than urging people to do their best. When we look at the results of our system there are two specific remarks which lead to question the goal setting that was used in the system. These remarks were: "It did give a plan, but it is questionable whether it was a good plan." and "Having an automatic plan is not pleasant (without adding your own preferences)". The goal is set based on the assessment week, in which participants only carried the activity monitor without getting feedback. The system

then does a proposal for a plan (and therefore thus the end goal and sub goals). After three weeks the user has the possibility to adjust this goal (our participants only used the system for a week). Participants do not find it completely clear how this proposal is composed and question how 'good' this plan is. Of course, the stage of behaviour change in which participants are will also influence their opinion about the plan, but the second comment given by participants gives an indication of what should be different (more influence/input from users).

When we look at important moderators of goal effect we see that goal commitment is one of the important moderators. Goal commitment will be influenced by the stage of behaviour change the user is in. Being in the wrong stage will influence goal commitment negatively. Secondly, appropriate feedback is an important moderator. In this case feedback is given in the form of a percentage of activity, while the goal of that day is getting a 100% activity score. Other feedback is given in the form of a weekly update on the average amount of activity the user has done the previous week. There is one big disadvantage with these forms in our opinion. Each day and each week, the goal is to get to a 100%, however this does not give insight in the progress since you started (the amount of activity that need to be done to get to 100% increases over time). We think that apart from giving insight in the goal of that day or week, it also is important to have a reference towards the overall goal, as was also mentioned in section 3.1.1. The last moderator is task complexity, which is closely related to the formulated plan and goal. When a plan is not good for the user, the level of task complexity might also be wrong. The formulated plan now automatically increases the amount of activity gradually over time. It might be that this results in the correct task complexity for the user, however based on the comments given above it seems more likely that task complexity was too low or too high (this we cannot say based on these comments, however based on other data like the amount of sports the participants do, task complexity was likely too low).

Elements of a behaviour change support system In previous sections we already discussed the importance of setting 'good' goals, and giving appropriate feedback towards this goal. Another important element of a behaviour change support system is to provide information about the benefits of the behaviour change. This helps to make an informed decision about the behaviour change, but also can serve as extra motivation for the behaviour change. We would like to argue that the SmarcoS-system could be improved in this area. Right now it does not give additional information apart from the progress of that day or week, towards the goal of that day or week. Providing more specific information about the benefits of being more active, combined with tips on how to be more active will help. Examples of more specific messages and tips are already discussed.

A behaviour change support system should reduce barriers, and increase self-efficacy of the user. We would like to argue that the SmarcoS-system does this partly. It can help users setting goals and help them attain those goals, but because it lacks on input that the user can give himself we argue that this could be improved. By better personalizing the goals, and feedback to the preferences of the user barriers are more effectively reduced, and self-efficacy is increased more.

In the theory we saw there are several important concepts that help create a successful behaviour change support system. We will shortly discuss these six concepts and discuss how well the SmarcoS system uses these. First of all, the authority role. The SmarcoS system has a kind of authorative role of coach which is good, because it give the system more power. But with this role expectations arise. Users expect to have a 'good' goal, and they expect to have 'good' feedback in the way you would expect this from a real coach. When expectations based on this role are on fulfilled this will influence attitude towards the system more, than when the system did not have this

role. Therefore, we would like to argue that is this case the authorative role lead to a more negative attitude towards the system. Secondly, messages should be personalised to the user's interests and characteristics. This is now only done by addressing the user with his or her own name. Apart from that the system is not capable of personalising messages. Based on the theory and results we argue that the system can be improved by personalising messages more. Thirdly, according to theory when a system has a character that looks like a person that communicates the messages, people are more likely to cooperate with the message than when it is communicates by a clearly unreal computer character. We did not find a difference between the two office worker versions (ECA vs. text). This could be caused by two things, either the character was human like enough, or, and this is more likely, other factors like the functioning of the system were more of influence than the difference in modality. Fourthly, accessibility on multiple devices increase effectivity. Right now we have the SmarcoS-diabetic version which has a smartphone and computer application, while this could have influence effectivity, it is not so likely. We would like to argue that accessibility on multiple devices will increase effectivity when the activity data is always up-to-date. Because users now have to dock the activity monitor, in order for the system to upload the data, there is a bias in the use of devices. The DirectLife website, strictly speaking is not part of the system, although participants of the SmarcoS-office worker evaluation did experience it this way. Therefore, accessibility on multiple devices is not implemented best as possible. Fifthly, an attractive system has greater persuasive power. When we look at the attractiveness of the system in the Attrakdiff questionnaires we see that the word pair of ugly–attractive scores relatively high for all versions of the system, indicating participants thought the system to be attractive. However, when we look at the whole attractiveness dimension of Attrakdiff we see that improvements can be made (mostly for the office worker versions). Therefore, although attractiveness is not bad, this can be improved. Finally, when a system is easy to use, it increases the likelihood that user will use the system. When we look at the technology acceptance construct Perceived ease of use, we see a scoring just below neutral. This means that the system could be easier to use, then it is now. Therefore, improvement can also be made here.

8 Conclusion

In this study we looked at the SmarcoS system, a behaviour change support system. We evaluated three different versions of this system, with two different target groups: diabetics and office workers. In this conclusion we will first give conclusions about the SmarcoS-diabetics evaluation, after which we will provide conclusions about the SmarcoS-office worker evaluation. Finally, we will give conclusions about the complete SmarcoS-system and implications for behaviour change support systems.

SmarcoS-diabetics evaluation While participants of the SmarcoS-diabetics evaluation were relatively positive about the system, it is not clear if they would like to use the system themselves, although they see benefits in such a system. This is partially due to the fact that all participants were diagnosed at least 1.5 years ago with diabetics, and all participants had a clear routine in taking their medication.

Based on the diaries and end-interview functional improvements were suggested. Pragmatic quality scores lowest in the Attrakdiff questionnaire, which is in line with the majority of comments about functional improvements. Usability can be improved according to the results of the system usability scale, in which "I found the system very cumbersome to use" and "I found the system unnecessarily complex" scored low, showing the most need for improvements.

The characteristics of the pill dispenser were not convenient for all users. Participants also thought docking of the activity monitor should be omitted. Reasons for this were, among others, because it will improve the feedback and because your activity data is always known and up-to-date. Feedback can now only be given afterwards. This was seen by the participants as unsatisfactory. Participants had a preference for the smartphone application or the computer application based on what they were most used to. People who lived with their smartphone preferred the smartphone application, while people who were less tied to their smartphone preferred to use the computer application.

Medication reminders should be more related to mealtimes instead of fixed clock times. Medication is often taken with each meal, but mealtimes can vary. Current implementation leads to unwanted situations. All feedback messages should be send at suitable moments, which participants considered to be in the morning, around lunch time or for example if they are inactive for too long (they did not specify too long). When messages are sent, whether these are medication messages or activity messages, the content of the message should be relevant, situation specific, concrete, and with tips.

Privacy is considered to be important, but no participants had objections to the gathered activity data. Most participants (3/5) also had no objection to gathering location data. This illustrates that people are willing to give up some of their privacy if they see the need for it or benefit enough from it. But also shows a boundary concerning private information. Location is seen as more personal than activity data or medication intake.

SmarcoS-office worker evaluation Participants of the SmarcoS-office worker evaluation had low expectations beforehand and were somewhat hesitant towards the system. The system did not meet the expectations of the participants, as was also seen in performance expectancy.

Participants would not like to use the system as it currently is. This is seen in the system usability scale, in the 'intention to use' construct, in the usefulness construct, and in the interview. The attitude towards the system also scores below average, indicating

that the attitude towards the system is not good. However, participants see potential value in the system as is illustrated by the fact that most participants think such a system could help people live a more healthy and balanced life, but only if the users want to change. It is debatable to which extend participants wanted to be more active, and were intrinsic motivated.

Most reported errors concerned functionality, but also content of messages was seen as wrong. User-friendliness should be improved, at least by improving the loading speed of the DirectLife website, which also has influenced the use of the system. Installation of the application should be easier (especially for the ECA version), which is also seen in perceived ease of use. The timing of the messages was not good and need to be preference based, and therefore more personalized. Participants thought the content of the messages was standard, administrative, not motivating, not diverse enough and there were not enough different types of messages. Content can be improved by making it more specific, concrete, personal and preference based. The construct of adaptability also shows that the system should be more adaptable to the needs of users.

The quality of coaching was not good according to participants. The content is bad, and some participants questioned the activity goal. Source credibility was mediocre, while trust scored just below average, indicating that participants do not wholly trust advice the system gives them and will probably not follow it. Therefore, this needs improvement. The construct facilitating conditions scores average, which tells us that participants had the resources and knowledge that was necessary to use the system. Anxiety scores well below average. This is good because you don't want users to be axious towards the system.

The complete SmarcoS-system Looking at our results in relation to the theory we see that the way goals are set in system might have been of negative influence, since there is no possibility for input of the users. Secondly when we look at the given feedback we see that this feedback is only given in relation to goal of the day or week, but not in relation to the overall goal. Giving feedback in relation to the end goal is important for motivational reasons.

When we look at the results of all evaluations there are notable differences. Because of the differences between all versions it is hard to compare them. The different versions of the system do no function the same, but because of the shared basis (see section 5.1.1) we can make a careful comparison.

We find difference in willingness to use the system, as well as in Attrakdiff scores, system usability scores, and privacy issues. All these differences can be carefully explained by the following: There is a difference in target groups, and therefore a difference in users. The main difference between those two groups is that diabetics have a disease and therefore can be considered patients. This implies that they are more aware of the importance of living a healthy and balanced life, which makes it likely that they are more willing to change and more likely to be in the correct stage of behaviour change (fourth stage: 'action'). It also influences the willingness to give up privacy in order for a system that can be beneficial to work properly.

Secondly, the emphasis of the system lies at different points. The SmarcoS-office worker system focused on becoming more active, while the SmarcoS-diabetics system also monitors medication intake. Medication intake probably has more direct effects which influences the focus of users more on this part of the system.

Finally, functional suggestions come back in both evaluations. Both evaluations bring forward a problem with docking of the activity monitor, mostly because participants realize that feedback can now only be given afterwards. This is seen as a problem and participants therefore would prefer automatic synchronization. All evaluations also bring forward the need to improve the content of messages. Messages should be improved by making them more specific, concrete, personal and preference based.

9 Future work

In this section we will give recommendation for future work based on the results of both evaluations, and the implications. We will first look at the evaluation itself.

Evaluation During the first evaluation of the SmarcoS-diabetic system diaries were given from day one. We would like to argue that it is better to first let the participants experience and use the system before asking them questions about the system. When you still wanted you participants to answer a diary every day (and keep the same diaries) it would be better to start with what was now diary three. This focused on where participants got their medical/health information from. Because it is unrelated to the system, you can ask participants to answer this on day one. This way participants get to experience the system more before having to answer questions about it.

Secondly, we question the appropriateness of the used smilies in the SmarcoS-diabetic evaluation, because they can be interpreted in different ways. When asked participants to place smilies to getter better insight in their opinion about the system, it is important that those smilies mean the same to everyone. We also find it debatable to which extend you can use the used words that had to be placed with pictures of the system. By using different fonts the connotation of the word can change. By using a cheery font, the word becomes more positive. Although this probably is of limited influence, it is easy to avoid by using only one font.

Thirdly, we think using diaries can be useful, when you use them in a different way. Now the diaries were a way to present de participants with a questionnaire each day. While it makes more sense to use diaries for writing down your experiences of that day with the system. Secondly, by using paper dairies it is likely that not questions are answered. This was not checked when collecting the dairies, which is a moment you can ask the participants to complete the missing questions. Now, some 'diaries' were not complete, this is a problem because you data which is already limited due to the number of participants will be even smaller.

Finally, we think it is important to make sure what you ask from you participants is not too childish. Using stickers, and other similar things, will not be appreciated by all participants. We also think it is important to make you 'diary' presentable, but not childish. This can give the impression to participants that you do not take them seriously, hindering their cooperation with the evaluation.

For the evaluation of the SmarcoS-office worker system, there are two questions we would have asked differently. The first is the open-end question concerning the expectations of the system. In further research we would suggest to divide the expectation question in separate questions: 'What do you expect the system will do?' and 'What do you expect the system will do for you?'. This way you can more easily distinguish between expectations concerning the system and expectations what the system would do for the user. Now we just now that all participants have relatively low expectations, concerning the system and what it will do for them. By making this division, it is easier to distinguish between reasons why expectations are low. Secondly, it would be better to rephrase the question about reasons why participants would want to be more active. Now it states: "Kun je aangeven wat redenen voor je zouden zijn op dit moment om je bewegingspatroon aan te passen?" (Could you provide possible reasons for you to adjust your activity pattern at this moment?) (see appendix C.1). Based on answers on this question and the interview, we are under the impression that this question was answered in a different way than we intended. It seemed that participants answered this question

as which reasons *could* now possibly be any valid reason for them to be more active. This is illustrated by the fact that participants who indicated to be very active (and also indicated that they did not want to be more active) still answered this question with reasons. However, we wanted to know whether they now wanted to change their activity pattern, and what the reasons were behind this.

In all evaluations the end-interview was the main source of data. When you would have a larger group of participants this would be inconvenient. Questionnaires then become more important. In further research we would suggest looking at which the information we now got from the interviews was the most important, and add questions to the questionnaires asking for this information.

Functionality Adjusting timing and the number of message with the system as it is right now might be difficult, because feedback can only be given accurately and relevant if the user has just docked their activity monitor. Part of the reason why messages are not seen as being relevant has to do with this problem, as also was noticed by multiple participants. This can partly be improved by making better use of the stored activity data. Better messages content wise and timings wise can be sent when making better use of reoccurring patterns in the activity data. For example around lunchtime there probably will be a sudden rise in activity due to someone getting away from their desk to eat lunch. This normally would occur each day at more or less the same time. The system should make good use of this data and send messages at the time this normally would occur, even if accurate data is not available. Of course this can be improved even further if the wish of participants was fulfilled and docking the activity monitor would be omitted, and the activity monitor would synchronize automatically.

Omitting the need to dock and synchronizing the activity monitor automatically removes an action from the user, which is important for improving usability, user experience, and user-friendliness. Secondly, the system then knows all activity data at any moment it needs it. It can have accurate data, and timing, content and the number of messages sent could be improved drastically. Having up-to-date data will make it easier to give accurate and relevant feedback. Rules can be made to send personalized messages, instead of sending a message at the end of the day that the user was not active enough. Most users probably knew that already and probably will benefit more from receiving motivational reminder at suitable moments during the day. These suitable moments are different for each participant. The different needs between users also make it impossible to have the kind of simple and standard rules for timing the system now uses. Therefore, new rules have to be defined to suit always accessible activity data, and to better accommodate timing and content of messages the way users want this.

When we look at similar existing systems we see that each system has advantages and disadvantage. A clear disadvantage of the SmarcoS-system is the need to dock the activity in order to synchronize the activity data. When we look at the 'It's LiFe!' tool from [89] we see that the idea behind this system is quite similar. However, it uses Bluetooth to synchronize their activity monitor with the smartphone application every 15 minutes. The application then synchronizes via an internet connection with the server. This is better suited for the need the participants express. Another system is the system of Roessingh Research and Development [71, 70]. In this system the user carries a ProMove3D inertial sensor node1, attached with a belt on the patient's hip, that constantly measures current activity intensity and transmits the measurements to the Smartphone over Bluetooth. Although, this device also has disadvantage (size of the monitor, and the belt), it illustrates that there are other possibilities when we look at monitoring the activity. It also shows that other research groups are focussing on the same kind of system. We think it would be very interesting for all groups to exchange information about their systems, or collaborate. This way, you can learn from other

researchers, you do not all have discover the same problems and solutions. Each system has its own advantages and disadvantage, combining the strong points of each system, will lead to a better system for all researchers. And after having a basic good working system, they all can focus on their own objectives.

Some participants questioned the plan the system generates and the goal that goes with the generated plan. They also would like to have more input in the goal, and the way they want to achieve that goal. It might be interesting to investigate how the way the plan/goal is generated/set influences the results. Secondly, it would be better to give users more insight in the reasoning behind the plan. By informing the users, they are more capable of making an informed decision, which will have a positive influence. It will increases the likelihood of certain behaviour and makes people more confident about the behaviour change. It also makes the goal behaviour seem more achievable. When behaviour seems more achievable, self-efficacy of people is increased.

We can use the results of both evaluations in different ways depending on the goal you have. When we want to optimize this system for actual use, results show us what needs to be improved and in which way. One of the most important finding in that case is that the docking of the activity monitor should be omitted (in all versions of the system, since this was a results in all evaluations). Messages should be more specific, concrete, personal and preference based. For example: 'You are not active enough today. To be more active you can go for a walk. When you take a 30 minute walk, you will burn 222 calories. This brings you 30% closer to your target. After your walk you will be at 60% of your target!'. Proposed activities and timings should be adjusted to the preferences of the user.

Furthermore, the system can benefit from some expansion for the diabetics target group in the form of, for example, glucose monitoring and maybe adjusting the pill dispenser. For the office workers the daily activity pattern should be added to the smartphone application (all extra functionality for the office workers of course can also be used in the case of diabetics).

When we want to further investigate whether there is a difference between feedback given by an ECA or by text, it is first important to improve feedback given by the system. Users saw so many limitations in the systems that any (subtle) differences between the ECA and the text version probably did not show. Secondly, we would find it interesting to investigated how realistic the ECA has to be in order to have the benefits that go with the appearance. The theory says: when the system has a character that looks like a person that communicates a message people are more likely to cooperate with the message than when it is communicated by a clearly unreal computer character, even if they find this character appealing and likable [74]. It might be that the behaviour of this character is of more importance than the looks (this was suggested by one of the participants).

Bibliography

[1] Philips directlife, September 2012. *http://www.directlife.philips.com.*

[2] Pill dispenser (sensemedic), September 2012. *http://www.evalan.com.*

[3] Smarcos, September 2012. *http://www.smarcos-project.eu.*

[4] A. Bandura. *Social foundations of thought and action: A social cognitive theory*, volume 1. Prentice-Hall, 1986.

[5] A. Bandura and D. Cervone. Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. *Journal of Personality and Social Psychology*, 45(5):1017–1028, 1983.

[6] A. Bangor, P. T. Kortum, and J. T. Miller. An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, 24(6):574–594, 2008.

[7] C. Bartneck, T. Kanda, H. Ishiguro, and N. Hagita. My robotic doppelgänger - a critical look at the uncanny valley, 2009.

[8] C. Bartneck, T. Kanda, O. Mubin, and A. A. Mahmud. The perception of animacy and intelligence based on a robot's embodiment. *2007 7th IEEERAS International Conference on Humanoid Robots*, pages 300–305, 2007.

[9] C. Bartneck, D. Kulic, and E. Croft. Measuring the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Development*, 471:37–44, 2008.

[10] C. Bartneck, M. Verbunt, O. Mubin, and A. A. Mahmud. *To kill a mockingbird robot*, pages 81–87. ACM Press, 2007.

[11] T. Bickmore, D. Mauer, F. Crespo, and T. Brown. Persuasion, task interruption and health regimen adherence. *Persuasive Technology*, 4744(17):1–11, 2007.

[12] T. W. Bickmore and R. W. Picard. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction*, 12(2):293–327, 2005.

[13] N. Bolger, A. Davis, and E. Rafaeli. Diary methods: capturing life as it is lived. *Annual Review of Psychology*, 54(1):579–616, 2003.

[14] J. Brooke. Sus - a quick and dirty usability scale. *Usability evaluation in industry*, 189:189–194, 1996.

[15] J. Brug, A. Oenema, W. Kroeze, and H. Raat. The internet and nutrition education: challenges and opportunities. *European Journal of Clinical Nutrition*, 59 Suppl 1:S130–S137; discussion S138–S139, 2005.

[16] M. K. Campbell, I. Tessaro, B. DeVellis, S. Benedict, K. Kelsey, L. Belton, and A. Sanhueza. Effects of a tailored health promotion program for female blue-collar workers: health works for women. *Preventive Medicine*, 34(3):313–323, 2002.

[17] J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. H. Vilhjálmsson, and H. Yan. Embodiment in conversational interfaces: Rea. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, pages 520–527, New York, USA, 1999.

[18] J. Cassell, T. Bickmore, L. Campbell, H. Vilhjálmsson, and H. Yan. Conversation as a system framework: Designing embodied conversational agents. In *Embodied Conversational Agents*, pages 29–63. MIT Press, 2000.

[19] J. P. Chin, V. A. Diehl, and K. L. Norman. *Development of an instrument measuring user satisfaction of the human-computer interface*, volume 218, pages 213–218. ACM, 1988.

[20] S. Consolvo, D. W. McDonald, and J. A. Landay. Theory-driven design strategies for technologies that support behavior change in everyday life. *Proceedings of the 27th international conference on Human factors in computing systems CHI 09*, CHI 2009(6):405, 2009.

[21] S. Consolvo and M. Walker. Using the experience sampling method to evaluate ubicomp applications, 2003.

[22] J. Côté, J. Yardley, J. Rock, W. Sedgwick, and J. Baker. An exploratory examination of the coaching behavior scale for sport. *Avante Research Note*, 5(2):82–92, 1999.

[23] F. D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319–340, 1989.

[24] F. D. Davis. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *International Journal of ManMachine Studies*, 38(3):475–487, 1993.

[25] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw. User acceptance of computer technology: a comparison of two theoretical models. *Manage. Sci.*, 35(8):982–1003, Aug. 1989.

[26] B. de Ruyter, P. Saini, P. Markopoulos, and A. van Breemen. Assessing the effects of building social intelligence in a robotic interface for the home. *Interacting with Computers*, 17(5):522–541, 2005.

[27] M. Erez. Feedback: A necessary condition for the goal setting-performance relationship. *Journal of Applied Psychology*, 62(5):624–627, 1977.

[28] M. Erez and I. Zidon. Effects of goal acceptance on the relationship of goal setting and task performance. *Journal of Applied Psychology*, 69:69–78, 1984.

[29] M. Fishbein and I. Ajzen. *Belief, attitude, intention, and behavior: An introduction to theory and research*, volume 10. Addison-Wesley, 1975.

[30] B. J. Fogg. *Persuasive computers: perspectives and research directions*, volume 98, pages 225–232. ACM Press/Addison-Wesley Publishing Co., 1998.

[31] B. J. Fogg. *Persuasive Technology: Using Computers to Change What We Think and Do*, volume 5, chapter 5, pages 89–120. Morgan Kaufmann, 2003.

[32] B. J. Fogg and C. Nass. Silicon sycophants: the effects of computers that flatter. *International Journal of Human-Computer Studies*, 46(5):551–561, 1997.

[33] M. V. D. Haak, M. De Jong, and P. J. Schellens. Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour Information Technology*, 22(5):339–351, 2003.

[34] M. Hassenzahl. *The thing and I: understanding the relationship between user and product*, volume 3, pages 31–42. Kluwer Academic Publishers, 2003.

[35] M. Hassenzahl. The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction*, 19(4):319–349, 2004.

[36] M. Hassenzahl. User experience (ux): Towards an experiential perspective on product quality. *Media*, pages 11–15, 2008.

[37] M. Hassenzahl, A. Beu, and M. Burmester. Engineering joy. *IEEE Software*, 18(1):70–76, 2001.

[38] M. Hassenzahl, M. Burmester, and F. Koller. Attrakdiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität. *Computer*, pages 187–196, 2003.

[39] M. Hassenzahl and N. Tractinsky. User experience Ű a research agenda. *Behaviour Information Technology*, 25(2):91–97, 2006.

[40] M. Heerink, B. Kröse, V. Evers, and B. Wielinga. Assessing acceptance of assistive social agent technology by older adults: the almere model. *International Journal of Social Robotics*, 2(4):361–375, 2010.

[41] M. Heerink, B. Kröse, V. Evers, and B. Wielinga. Studying the acceptance of a robotic agent by elderly users. *International Journal of Assistive Robotics and Mechatronics*, 7(3):33–43, 2006.

[42] M. Heerink, B. Kröse, V. Evers, and B. Wielinga. Measuring acceptance of an assistive social robot: a suggested toolkit. *ROMAN 2009 The 18th IEEE International Symposium on Robot and Human Interactive Communication*, 7:528–533, 2009.

[43] A. R. Hendrickson, P. D. Massey, and T. P. Cronan. On the test-retest reliability of perceived usefulness and perceived ease of use scales. *MIS Quarterly*, 17(2):227–230, 1993.

[44] D. C. Huddy, J. L. Hebert, G. C. Hyner, and R. L. Johnson. Facilitating changes in exercise behavior: effect of structured statements of intention on perceived barriers to action. *Psychol Rep*, 76(3 Pt 1):867–75, 1995.

[45] Ergonomics of Human System Interaction, ISO 9241-11.

[46] A. D. Jacobs, A. S. Ammerman, S. T. Ennett, M. K. Campbell, K. W. Tawney, S. A. Aytur, S. W. Marshall, J. C. Will, and W. D. Rosamond. Effects of a tailored follow-up intervention on health behaviors, beliefs, and attitudes. *Journal of womens health 2002*, 13(5):557–568, 2004.

[47] M. Joosse, A. Sardar, and V. Evers. Behave: A set of measures to assess users' attitudinal and non-verbal behavioral responses to a robot's social behaviors. In B. Mutlu, C. Bartneck, J. Ham, V. Evers, and T. Kanda, editors, *Social Robotics*, volume 7072 of *Lecture Notes in Computer Science*, pages 84–94, Berlin, Germany, November 2011. Springer Verlag.

[48] D. Kahneman, A. B. Krueger, D. A. Schkade, N. Schwarz, and A. A. Stone. A survey method for characterizing daily life experience: the day reconstruction method. *Science*, 306(5702):1776–80, 2004.

[49] C.-M. Karat, R. Campbell, and T. Fiegel. *Comparison of empirical testing and walkthrough methods in user interface evaluation*, pages 397–404. ACM Press, 1992.

[50] S. Kiesler, L. Sproull, and K. Waters. A prisoner's dilemma experiment on cooperation with people and human-like computers. *Journal of Personality and Social Psychology*, 70(1):47–65, 1996.

[51] R. Klaassen, J. Hendrix, D. Reidsma, and R. Akker op den. Elckerlyc goes mobile enabling technology for ecas in mobile applications. In *UBICOMM*, pages 41–47, 2012.

[52] H. J. Klein, M. J. Wesson, J. R. Hollenbeck, and B. J. Alge. Goal commitment and the goal-setting process: conceptual clarification and empirical synthesis. *Journal of Applied Psychology*, 84(6):885–896, 1999.

[53] M. W. Kreuter and V. J. Strecher. Do tailored behavior change messages enhance the effectiveness of health risk appraisal? results from a randomized trial. *Health Education Research*, 11(1):97–105, 1996.

[54] D. Kulic and E. Croft. Anxiety detection during human-robot interaction. *2005 IEEERSJ International Conference on Intelligent Robots and Systems*, pages 616–621, 2005.

[55] D. Kulic and E. Croft. Estimating robot induced affective state using hidden markov models. *ROMAN 2006 The 15th IEEE International Symposium on Robot and Human Interactive Communication*, (i):257–262, 2006.

[56] R. E. LaPorte and R. Nath. Role of performance goals in prose learning. *Journal of Educational Psychology*, 68(3):260–264, 1976.

[57] K. M. Lee, N. Park, and H. Song. Can a robot be perceived as a developing creature? *Human Communication Research*, 31(4):538–563, 2005.

[58] H. Leventhal and L. Cameron. Behavioral theories and the problem of compliance. *Patient Education and Counseling*, 10(2):117–138, 1987.

[59] D. A. Lieberman. *Interactive video games for health promotion: Effects on knowledge, self-efficacy, social support, and health*, pages 103–120. Lawrence Erlbaum Associates, 1997.

[60] E. A. Locke, D.-O. Chah, S. Harrison, and N. Lustgarten. Separating the effects of goal specificity from goal level. *Organizational Behavior and Human Decision Processes*, 43(2):270–287, April 1989.

[61] E. A. Locke and G. P. Latham. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9):705–717, 2002.

[62] E. A. Locke and G. P. Latham. What should we do about motivation theory? six recommendations for the twenty-first century. *Academy of Management Review*, 29(3):388, 2004.

[63] J. C. McCroskey. Scales for the measurement of ethos. *Speech Monographs*, 33(1):65–72, 1966.

[64] J. C. McCroskey and T. J. Young. Ethos and credibility: The construct and its measurement after three decades. *Central States Speech Journal*, 32(1):24–34, 1981.

[65] J. L. Monahan. I don't know it but i like you the influence of non-conscious affect on person perception. *Human Communication Research*, 24(4):480–500, 1998.

[66] L. M. Neville, B. O?Hara, and A. Milat. Computer-tailored physical activity behavior change interventions targeting adults: a systematic review. *The international journal of behavioral nutrition and physical activity*, 6(1):30, 2009.

[67] S. M. Noar, C. N. Benac, and M. S. Harris. Does tailoring matter? meta-analytic review of tailored print health behavior change interventions. *Psychological Bulletin*, 133(4):673–693, 2007.

[68] G. J. Norman, M. F. Zabinski, M. A. Adams, D. E. Rosenberg, A. L. Yaroch, and A. A. Atienza. A review of ehealth interventions for physical activity and dietary behavior change. *American Journal of Preventive Medicine*, 33(4):336–345, 2007.

[69] H. Oinas-kukkonen. Behavior change support systems: A research model and agenda. *Persuasive Technology*, LNCS 6137:4–14, 2010.

[70] H. Op Den Akker, V. M. Jones, L. S. Moualed, and H. J. Hermens. *A Context-Aware Adaptive Feedback Agent for Activity Monitoring and Coaching*, pages 129–132. Noldus Information Technology bv, 2012.

[71] H. Op Den Akker, L. S. Moualed, V. M. Jones, and H. J. Hermens. A self-learning personalized feedback agent for motivating physical activity. *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies ISABEL 11*, pages 1–5, 2011.

[72] L. Oshlyansky, S. Park, P. Cairns, and H. Thimbleby. Validating the unified theory of acceptance and use of technology (utaut) tool cross-culturally. *Analysis*, 2(September):83–86, 2007.

[73] S. Parise, S. Kiesler, L. Sproull, and K. Waters. My partner is a real dog: cooperation with social agents. In *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, CSCW '96, pages 399–408, New York, NY, USA, 1996. ACM.

[74] S. Parise, S. Kiesler, L. Sproull, and K. Waters. Cooperating with life-like interface agents. *Computers in Human Behavior*, 15(2):123–142, 1999.

[75] K. Patrick, S. S. Intille, and M. F. Zabinski. An ecological framework for cancer communication: Implications for research. *Journal of Medical Internet Research*, 7(3):e23, 2005.

[76] H. Petrie and N. Bevan. The evaluation of accessibility , usability and user experience. *The Universal Access Handbook*, pages 299–315, 2009.

[77] M. Pontier and G. F. Siddiqui. A Virtual Therapist That Responds Empathically to Your Answers. In *Proceedings of the 8th international conference on Intelligent Virtual Agents*, IVA '08, pages 417–425, Berlin, Heidelberg, 2008. Springer-Verlag.

[78] A. Powers and S. Kiesler. *The advisor robot: tracing people's mental model from a robot's physical attributes*, volume 1, pages 218–225. ACM, 2006.

[79] J. O. Prochaska, C. C. DiClemente, and J. C. Norcross. In search of how people change. applications to addictive behaviors. *American Psychologist*, 47(9):1102–14, 1992.

[80] J. O. Prochaska and W. F. Velicer. The transtheoretical model of health behavior change. *American journal of health promotion AJHP*, 12(1):38–48, 1997.

[81] J. O. Prochaska, W. F. Velicer, J. S. Rossi, M. G. Goldstein, B. H. Marcus, W. Rakowski, C. Fiore, L. L. Harlow, C. A. Redding, and D. Rosenbloom. Stages of change and decisional balance for 12 problem behaviors. *Health psychology official journal of the Division of Health Psychology American Psychological Association*, 13(1):39–46, 1994.

[82] D. Reidsma and H. v. Welbergen. Elckerlyc in practice on the integration of a bml realizer in real applications. In *Proc. of Intetain 2011*, 2011.

[83] D. Schulman and T. Bickmore. Persuading users through counseling dialogue with a conversational agent. *Proceedings of the 4th International Conference on Persuasive Technology Persuasive 09*, page 1, 2009.

[84] G. Solano Méndez and D. Reidsma. A bml based embodied conversational agent for a personality detection program. In H. H. Vilhjálmsson, S. Kopp, S. Marsella, and K. R. Thórisson, editors, *Intelligent Virtual Agents*, volume 6895 of *Lecture Notes in Computer Science*, pages 468–469, Berlin, Germany, September 2011. Springer Verlag.

[85] B. Sternthal, R. Dholakia, and C. Leavitt. The persuasive effect of source credibility: Tests of cognitive response. *Journal of Consumer Research*, 4(4):252–260, 1978.

[86] P. Touset. Natural Language Processing for a Psychometer. Master's thesis, Human Media Interaction, University of Twente, 2007.

[87] T. Tullis, S. Fleischman, M. McNulty, C. Cianchette, and M. Bergel. *An Empirical Comparison of Lab and Remote Usability Testing of Web Sites*. UPA, 2002.

[88] T. S. Tullis and J. N. Stetson. A comparison of questionnaires for assessing website usability. *Usability Professional Association Conference*, pages 1–12, 2004.

[89] S. van der Weegen, R. Verwey, M. Spreeuwenberg, H. Tange, T. van der Weijden, and L. de Witte. It's life!: A monitoring-and feedback tool to stimulate physical activity, embedded in primary care. In *Measuring Behavior*.

[90] V. Venkatesh and F. D. Davis. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2):186–204, 2000.

[91] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis. User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3):425–478, 2003.

[92] M. Vervloet, L. van Dijk, J. Santen-Reestman, B. van Vlijmen, M. L. Bouvy, and D. H. de Bakker. Improving medication adherence in diabetes type 2 patients through real time medication monitoring: a randomised controlled trial to evaluate the effect of monitoring patientsŠmedication use combined with short message service (sms) reminders. *BMC Health Services Research*, 11(1):5–8, 2011.

[93] R. M. Warner and D. B. Sugarman. Attributions of personality based on physical appearance, speech, and handwriting. *Journal of Personality and Social Psychology*, 50(4):792–799, 1986.

A Questionnaires

A.1 System Usability Scale

For each of the following statements, mark one circle that best describes your reactions to the system you used.

Rate your immediate response to each item, rather than thinking about items for a long time.

All items should be checked. If you feel that you cannot respond to a particular item, you should mark the centre point of the scale.

		Strongly disagree				**Strongly agree**
1)	I think that I would like to use this system frequently	○	○	○	○	○
2)	I found the system unnecessarily complex	○	○	○	○	○
3)	I thought the system was easy to use	○	○	○	○	○
4)	I think that I would need the support of a technical person to be able to use this system	○	○	○	○	○
5)	I found the various functions in this system were well integrated	○	○	○	○	○
6)	I thought there was too much inconsistency in this system	○	○	○	○	○
7)	I would imagine that most people would learn to use this system very quickly	○	○	○	○	○
8)	I found the system very cumbersome to use	○	○	○	○	○
9)	I felt very confident using the system	○	○	○	○	○
10)	I needed to learn a lot of things before I could get going with this system	○	○	○	○	○

Scoring SUS To calculate the SUS score, first sum the score contributions from each item. Each item's score contribution will range from 0 to 4. For items 1, 3, 5, 7, and 9 the score contribution is the scale position minus 1. For items 2, 4, 6, 8 and 10, the contribution is 5 minus the scale position. Multiply the sum of the scores by 2.5 to obtain the overall value of SUS. SUS scores have a range of 0 to 100. The SUS is not a percentage!

Source: [14]

A.2 Questionnaire for User Interface Satisfaction

User Evaluation of an Interactive Computer System. For each of the following questions, fill in 0-9 or leave blank if question is not applicable.

Overall reaction to the software											
1)		terrible	○	○	○	○	○	○	○	○	○	○	wonderful
2)		difficult	○	○	○	○	○	○	○	○	○	○	easy
3)		frustrating	○	○	○	○	○	○	○	○	○	○	satisfying
4)		inadequate power	○	○	○	○	○	○	○	○	○	○	adequate power
5)		dull	○	○	○	○	○	○	○	○	○	○	stimulating
6)		rigid	○	○	○	○	○	○	○	○	○	○	flexible
Screen											
7)	Reading characters on the screen	hard	○	○	○	○	○	○	○	○	○	○	easy
8)	Highlighting simplifies task	not at all	○	○	○	○	○	○	○	○	○	○	very much
9)	Organization of information	confusing	○	○	○	○	○	○	○	○	○	○	very clear
10)	Sequence of screens	confusing	○	○	○	○	○	○	○	○	○	○	very clear
Terminology and system information											
11)	Use of terms throughout system	inconsistent	○	○	○	○	○	○	○	○	○	○	consistent
12)	Terminology related to task	never	○	○	○	○	○	○	○	○	○	○	always
13)	Position of messages on screen	inconsistent	○	○	○	○	○	○	○	○	○	○	consistent
14)	Prompts for input	confusing	○	○	○	○	○	○	○	○	○	○	clear
15)	Computer informs about its progress	never	○	○	○	○	○	○	○	○	○	○	always
16)	Error messages	unhelpful	○	○	○	○	○	○	○	○	○	○	helpful

Continute on the next page.

Source:[19]

Learning											
17)	Learning to operate the system	difficult ○	○	○	○	○	○	○	○	○	easy ○
18)	Exploring new features by trial and error	difficult ○	○	○	○	○	○	○	○	○	easy ○
19)	Remembering names and use of commands	difficult ○	○	○	○	○	○	○	○	○	easy ○
20)	Performing tasks in straightforward	never ○	○	○	○	○	○	○	○	○	always ○
21)	Help messages on the screen	unhelpful ○	○	○	○	○	○	○	○	○	helpful ○
22)	Supplemental reference materials	confusing ○	○	○	○	○	○	○	○	○	clear ○
System capabilities											
23)	System speed	too slow ○	○	○	○	○	○	○	○	○	fast enough ○
24)	System reliability	unreliable ○	○	○	○	○	○	○	○	○	reliable ○
25)	System tends to be	noisy ○	○	○	○	○	○	○	○	○	quite ○
26)	Correctin your mistakes	difficult ○	○	○	○	○	○	○	○	○	easy ○
27)	Designed for all levels of users	never ○	○	○	○	○	○	○	○	○	always ○

A.3 User experience (AttrakDiff2)

Following, are pairs of words to assist you in your evaluation. Each pair represents extreme contrasts. The possibilities between the extremes enable you to describe the intensity of the quality you choose.

An example:

disagreeable ○ ○ ○ ○ ● ○ ○ likeable

This evaluation tells us that the product is predominantly likable, but that there is marginal room for improvement.

Do not spend time thinking about the word-pairs. Try to give a spontaneous response. You may feel that some pairs of terms do not adequately describe the product. In this case please still be sure to give an answer. Keep in mind that there is no right or wrong answer. Your personal opinion is what counts!

With the help of the word-pairs please enter what you consider the most appropriate description for the product. Please mark your choice in every line!

PQ*	technical	○	○	○	○	○	○	○	human
	complicated	○	○	○	○	○	○	○	simple
	impractical	○	○	○	○	○	○	○	practical
	cumbersome	○	○	○	○	○	○	○	straightforward
	unpredictable	○	○	○	○	○	○	○	predictable
	confusing	○	○	○	○	○	○	○	clearly structured
	unruly	○	○	○	○	○	○	○	manageable
HQ-I*	isolating	○	○	○	○	○	○	○	connective
	unprofessional	○	○	○	○	○	○	○	professional
	tacky	○	○	○	○	○	○	○	stylish
	cheap	○	○	○	○	○	○	○	premium
	alienating	○	○	○	○	○	○	○	integrating
	separates me	○	○	○	○	○	○	○	brings me closer
	unpresentable	○	○	○	○	○	○	○	presentable
HQ-S*	conventional	○	○	○	○	○	○	○	inventive
	unimaginative	○	○	○	○	○	○	○	creative
	cautious	○	○	○	○	○	○	○	bold
	conservative	○	○	○	○	○	○	○	innovative
	dull	○	○	○	○	○	○	○	captivating
	undemanding	○	○	○	○	○	○	○	challenging
	ordinary	○	○	○	○	○	○	○	novel
ATT*	unpleasant	○	○	○	○	○	○	○	pleasant
	ugly	○	○	○	○	○	○	○	attractive
	disagreeable	○	○	○	○	○	○	○	likeable
	rejecting	○	○	○	○	○	○	○	inviting
	bad	○	○	○	○	○	○	○	good
	repelling	○	○	○	○	○	○	○	appealing
	discouraging	○	○	○	○	○	○	○	motivating

* PQ = Pragmatic quality, HQ-I = Hedonic quality - Identity, HQ-S = Hedonic quality - Stimulation, ATT = Attractiveness

Source:[38]

A.4 User Acceptance

A.4.1 Technology Acceptance Model

		extremely	quite	slightly	neither	slightly	quite	extremely
Perceived usefulness		likely						unlikely
1)	Using the system in my job would enable me to accomplish tasks more quickly	○	○	○	○	○	○	○
2)	Using the system would improve my job performance	○	○	○	○	○	○	○
3)	Using the system in my job would increase my productivity	○	○	○	○	○	○	○
4)	Using the system would enhance my effectiveness on the job	○	○	○	○	○	○	○
5)	Using the system would make it easier to do my job	○	○	○	○	○	○	○
6)	I would find the system useful in my job	○	○	○	○	○	○	○
Perceived ease of use		likely						unlikely
7)	Learning to operate the system would be easy for me	○	○	○	○	○	○	○
8)	I would find it easy to get the system to do what I want it to do	○	○	○	○	○	○	○
9)	My interaction with the system would be clear and understandable	○	○	○	○	○	○	○
10)	I would find the system to be flexible to interact with	○	○	○	○	○	○	○
11)	It would be easy for me to become skilful at using the system	○	○	○	○	○	○	○
12)	I would find the system easy to use	○	○	○	○	○	○	○

Source:[23]

A.4.2 Technology Acceptance Model2

Continute on the next page.

		strongly disagree	moderately disagree	somewhat disagree	neutral	somewhat agree	moderately agree	strongly agree
Performance expectancy								
1)	Assuming I have access to the system, I intend to use it	○	○	○	○	○	○	○
2)	Given that I have access to the system, I predict that I would use it	○	○	○	○	○	○	○
Perceived usefulness								
3)	Using the system improves my performance in my job	○	○	○	○	○	○	○
4)	Using the system in my job increases my productivity	○	○	○	○	○	○	○
5)	Using the system enhances my effectiveness in my job	○	○	○	○	○	○	○
6)	I find the system to be useful in my job	○	○	○	○	○	○	○
Perceived ease of use								
7)	My interaction with the system is clear and understandable	○	○	○	○	○	○	○
8)	Interacting with the system does not require a lot of my mental effort	○	○	○	○	○	○	○
9)	I find the system to be easy to use.	○	○	○	○	○	○	○
10)	I find it easy to get the system to do what I want it to do	○	○	○	○	○	○	○
Subjective Norm								
11)	People who influence my behaviour think that I should use the system	○	○	○	○	○	○	○
12)	People who are important to me think that I should use the system	○	○	○	○	○	○	○

Source:[90]

		strongly disagree	moderately disagree	somewhat disagree	neutral	somewhat agree	moderately agree	strongly agree
Voluntariness								
13)	My use of the system is voluntary.	○	○	○	○	○	○	○
14)	My supervisor does not require me to use the system.	○	○	○	○	○	○	○
15)	Although it might be helpful, using the system is certainly not compulsory in my job.	○	○	○	○	○	○	○
Image								
16)	People in my organization who use the system have more prestige than those who do not.	○	○	○	○	○	○	○
17)	People in my organization who use the system have a high profile	○	○	○	○	○	○	○
18)	Having the system is a status symbol in my organization	○	○	○	○	○	○	○
Job Relevance								
19)	In my job, usage of the system is important	○	○	○	○	○	○	○
20)	In my job, usage of the system is relevant	○	○	○	○	○	○	○
Output Quality								
21)	The quality of the output I get from the system is high	○	○	○	○	○	○	○
22)	I have no problem with the quality of the system's output	○	○	○	○	○	○	○
Result Demonstrability								
23)	I have no difficulty telling others about the results of using the system	○	○	○	○	○	○	○
24)	I believe I could communicate to others the consequences of using the system.	○	○	○	○	○	○	○
25)	The results of using the system are apparent to me	○	○	○	○	○	○	○
26)	I would have difficulty explaining why using the system may or may not be beneficial	○	○	○	○	○	○	○

A.4.3 Unified Theory of Acceptance and Use of Technology

Performance expectancy								
1)	I would find the system useful in my job	○	○	○	○	○	○	○
2)	Using the system enables me to accomplish tasks more quickly	○	○	○	○	○	○	○
3)	Using the system increases my productivity	○	○	○	○	○	○	○
4)	If I use the system, I will increase my chances of getting a raise	○	○	○	○	○	○	○
Effort expectancy								
5)	My interaction with the system would be clear and understandable	○	○	○	○	○	○	○
6)	It would be easy for me to become skilful at using the system	○	○	○	○	○	○	○
7)	I would find the system easy to use	○	○	○	○	○	○	○
8)	Learning to operate the system is easy for me	○	○	○	○	○	○	○
Attitude toward using technology								
9)	Using the system is a bad/good idea.	○	○	○	○	○	○	○
10)	The system makes work more interesting	○	○	○	○	○	○	○
11)	Working with the system is fun	○	○	○	○	○	○	○
12)	I like working with the system	○	○	○	○	○	○	○
Social influence								
13)	People who influence my behaviour think that I should use the system	○	○	○	○	○	○	○
14)	People who are important to me think that I should use the system	○	○	○	○	○	○	○
15)	The senior management of this business has been helpful in the use of the system	○	○	○	○	○	○	○
16)	In general, the organization has supported the use of the system	○	○	○	○	○	○	○

Continute on the next page.

Source:[91]

Facilitating conditions								
17)	I have the resources necessary to use the system	○	○	○	○	○	○	○
18)	I have the knowledge necessary to use the system	○	○	○	○	○	○	○
19)	The system is not compatible with other systems I use	○	○	○	○	○	○	○
20)	A specific person (or group) is available for assistance with system difficulties	○	○	○	○	○	○	○
Self-efficacy								
21)	I could complete a job or task using the system If there was no one around to tell me what to do as I go	○	○	○	○	○	○	○
22)	I could complete a job or task using the system If I could call someone for help if I got stuck	○	○	○	○	○	○	○
23)	I could complete a job or task using the system If I had a lot of time to complete the job for which the software was provided	○	○	○	○	○	○	○
24)	I could complete a job or task using the system If I had just the built-in help facility for assistance	○	○	○	○	○	○	○
Anxiety								
25)	I feel apprehensive about using the system	○	○	○	○	○	○	○
26)	It scares me to think that I could lose a lot of information using the system by hitting the wrong key	○	○	○	○	○	○	○
27)	I hesitate to use the system for fear of making mistakes I cannot correct	○	○	○	○	○	○	○
28)	The system is somewhat intimidating to me	○	○	○	○	○	○	○
Behavioural intention to use the system								
29)	I intend to use the system in the next <n> months	○	○	○	○	○	○	○
30)	I predict I would use the system in the next <n> months	○	○	○	○	○	○	○
31)	I plan to use the system in the next <n> months	○	○	○	○	○	○	○

A.5 Robot acceptance

A.5.1 Heerink

This questionnaire was presented by an interviewer, since some participants in this study had trouble reading.

		absolutely not	not	neutral	yes	absolutely yes
Computer experience						
1)	Have you ever used a computer?					
2)	Do you still sometimes use a computer?					
Conversational acceptance						
3)	Did you feel uncomfortable talking to a robot?	○	○	○	○	○
Performance expectancy						
4)	Do you think iCat would be useful to you?	○	○	○	○	○
5)	Do you think iCat would help you do things?	○	○	○	○	○
Effort expectancy						
6)	As you have noticed, you control iCat by speech. Do you think you can easily communicate with it that way?	○	○	○	○	○
7)	Do you think you can quickly learn how to control iCat?	○	○	○	○	○
8)	Do you think iCat is easy to use?	○	○	○	○	○
Social Influence						
9)	Do you think many people would find it nice if you would have an iCat?	○	○	○	○	○
10)	Are these people whose opinion you value?	○	○	○	○	○
11)	Are these people who are important to you?	○	○	○	○	○
12)	Do you think the staff would find it nice if you would have an iCat?	○	○	○	○	○

Continute on the next page.

Source:[41]

		absolutely not	not	neutral	yes	absolutely yes
Social Abilities						
13)	Did you ding iCat a pleasant conversational partner?	○	○	○	○	○
14)	Would you consider iCat to be social?	○	○	○	○	○
15)	Would you trust iCat if it gave you advice?	○	○	○	○	○
16)	Would you follow iCat's advice?	○	○	○	○	○
17)	Do you feel understood by iCat?	○	○	○	○	○
Attitude toward using technology						
18)	Do you think it is a good idea to use iCat?	○	○	○	○	○
19)	Would you like to use iCat?	○	○	○	○	○
Self-efficacy						
20)	Do you think you could work with iCat without any help?	○	○	○	○	○
21)	Do you think you could work with iCat if you could call someone for help?	○	○	○	○	○
22)	Do you think you could work with iCat if you had a good manual?	○	○	○	○	○
Anxiety						
23)	Do you feel at ease with iCat?	○	○	○	○	○
24)	If you were to use iCat, would you be afraid to make mistakes or break something?	○	○	○	○	○
Intention to use						
25)	If you could have iCat, would you want it immediately?	○	○	○	○	○
26)	If you could have iCat, would you want it in a view months?	○	○	○	○	○
27)	If you could have iCat, would you want it in a few years?	○	○	○	○	○

A.5.2 Almere model

		strongly disagree	disagree	neutral	agree	strongly agree
Anxiety						
1)	If I should use the robot, I would be afraid to make mistakes with it	○	○	○	○	○
2)	If I should use the robot, I would be afraid to break something	○	○	○	○	○
3)	I find the robot scary	○	○	○	○	○
4)	I find the robot intimidating	○	○	○	○	○
Attitude						
5)	I think it's a good idea to use the robot	○	○	○	○	○
6)	The robot would make my life more interesting	○	○	○	○	○
7)	It's good to make use of the robot	○	○	○	○	○
Facilitating conditions						
8)	I have everything I need to make good use of the robot.	○	○	○	○	○
9)	I know enough of the robot to make good use of it.	○	○	○	○	○
Intention to use						
10)	I think I'll use the robot during the next few days	○	○	○	○	○
11)	I am certain to use the robot during the next few days	○	○	○	○	○
12)	I'm planning to use the robot during the next few days	○	○	○	○	○
Perceived adaptability						
13)	I think the robot can be adaptive to what I need	○	○	○	○	○
14)	I think the robot will only do what I need at that particular moment	○	○	○	○	○
15)	I think the robot will help me when I consider it to be necessary	○	○	○	○	○
Perceived enjoyment						
16)	I enjoy the robot talking to me	○	○	○	○	○
17)	I enjoy doing things with the robot	○	○	○	○	○
18)	I find the robot enjoyable	○	○	○	○	○
19)	I find the robot fascinating	○	○	○	○	○
20)	I find the robot boring	○	○	○	○	○

Continute on the next page.

Source:[42]

		strongly disagree	disagree	neutral	agree	strongly agree
Perceived ease of use						
21)	I think I will know quickly how to use the robot	○	○	○	○	○
22)	I find the robot easy to use	○	○	○	○	○
23)	I think I can use the robot without any help	○	○	○	○	○
24)	I think I can use the robot when there is someone around to help me	○	○	○	○	○
25)	I think I can use the robot when I have a good manual	○	○	○	○	○
Perceived sociability						
26)	I consider the robot a pleasant conversational partner	○	○	○	○	○
27)	I find the robot pleasant to interact with	○	○	○	○	○
28)	I feel the robot understands me	○	○	○	○	○
29)	I think the robot is nice	○	○	○	○	○
Perceived usefulness						
30)	I think the robot is useful to me	○	○	○	○	○
31)	It would be convenient for me to have the robot	○	○	○	○	○
32)	I think the robot can help me with many things	○	○	○	○	○
Social influence						
33)	I think the staff would like me using the robot	○	○	○	○	○
34)	I think it would give a good impression if I should use the robot	○	○	○	○	○
Social presence						
35)	When interacting with the robot I felt like I'm talking to a real person	○	○	○	○	○
36)	It sometimes felt as if the robot was really looking at me	○	○	○	○	○
37)	I can imagine the robot to be a living creature	○	○	○	○	○
38)	I often think the robot is not a real person	○	○	○	○	○
39)	Sometimes the robot seems to have real feelings	○	○	○	○	○
Trust						
40)	I would trust the robot if it gave me advice	○	○	○	○	○
41)	I would follow the advice the robot gives me	○	○	○	○	○

A.5.3 GODSPEED

GODSPEED I: Anthropomorphism						
Please rate your impression of the robot on these scales :						
fake	○	○	○	○	○	natural
machinelike	○	○	○	○	○	humanlike
unconscious	○	○	○	○	○	conscious
artificial	○	○	○	○	○	lifelike
moving rigidly	○	○	○	○	○	moving elegantly
GODSPEED II: Animacy						
Please rate your impression of the robot on these scales :						
dead	○	○	○	○	○	alive
stagnant	○	○	○	○	○	lively
mechanical	○	○	○	○	○	organic
artificial	○	○	○	○	○	lifelike
inert	○	○	○	○	○	interactive
apathetic	○	○	○	○	○	responsive
GODSPEED III: Likeability						
Please rate your impression of the robot on these scales :						
dislike	○	○	○	○	○	like
unfriendly	○	○	○	○	○	friendly
unkind	○	○	○	○	○	kind
unpleasant	○	○	○	○	○	pleasant
awful	○	○	○	○	○	nice
GODSPEED IV: Perceived Intelligence						
Please rate your impression of the robot on these scales :						
incompetent	○	○	○	○	○	competent
ignorant	○	○	○	○	○	knowledgeable
irresponsible	○	○	○	○	○	responsible
unintelligent	○	○	○	○	○	intelligent
foolish	○	○	○	○	○	sensible
GODSPEED V: Perceived Safety						
Please rate your emotional state on these scales :						
anxious	○	○	○	○	○	relaxed
agitated	○	○	○	○	○	calm
quiescent	○	○	○	○	○	surprised

Source:[9]

A.6 Source credibility

A.6.1 Source credibility twelve item version

On the scales below, please indicate your feelings about..... Circle the number between the adjectives which best represents your feelings about..... Numbers "1" and "7" indicate a very strong feeling. Numbers "2" and "6" indicate a strong feeling. Numbers "3" and "5" indicate a fairly weak feeling. Number "4" indicates you are undecided or do not understand the adjectives themselves. Please work quickly. There are no right or wrong answers.

Authoritativeness								
reliable	1	2	3	4	5	6	7	unreliable *
uninformed	1	2	3	4	5	6	7	informed
unqualified	1	2	3	4	5	6	7	qualified
intelligent	1	2	3	4	5	6	7	unintelligent *
valuable	1	2	3	4	5	6	7	worthless *
inexpert	1	2	3	4	5	6	7	expert
Character								
honest	1	2	3	4	5	6	7	dishonest *
unfriendly	1	2	3	4	5	6	7	friendly
pleasant	1	2	3	4	5	6	7	unpleasant *
selfish	1	2	3	4	5	6	7	unselfish
awful	1	2	3	4	5	6	7	nice
virtuous	1	2	3	4	5	6	7	sinful *

Note. Items are presented here grouped by dimensions. Users should randomly order the bipolar adjectives to avoid response set error variance. Reverse scoring should be performed for items with asterisks.

Source:[63]

A.6.2 Source credibility fiveteen item version

On the scales below, please indicate your feelings about..... Circle the number between the adjectives which best represents your feelings about..... Numbers “1” and “7” indicate a very strong feeling. Numbers “2” and “6” indicate a strong feeling. Numbers “3” and “5” indicate a fairly weak feeling. Number “4” indicates you are undecided or do not understand the adjectives themselves. Please work quickly. There are no right or wrong answers.

Sociability								
good-natured	1	2	3	4	5	6	7	irritable *
cheerful	1	2	3	4	5	6	7	gloomy *
unfriendly	1	2	3	4	5	6	7	friendly
Extroversion								
timid	1	2	3	4	5	6	7	bold
verbal	1	2	3	4	5	6	7	quiet *
talkative	1	2	3	4	5	6	7	silent *
Competence								
expert	1	2	3	4	5	6	7	inexpert *
unintelligent	1	2	3	4	5	6	7	intelligent
intellectual	1	2	3	4	5	6	7	narrow *
Composure								
poised	1	2	3	4	5	6	7	nervous *
tense	1	2	3	4	5	6	7	relaxed
calm	1	2	3	4	5	6	7	anxious *
Character								
dishonest	1	2	3	4	5	6	7	honest
unsympathetic	1	2	3	4	5	6	7	sympathetic
good	1	2	3	4	5	6	7	bad *

Note. Items are presented here grouped by dimensions. Users should randomly order the bipolar adjectives to avoid response set error variance. Reverse scoring should be performed for items with asterisks.

Source:[63]

A.7 Coaching

A.7.1 Coaching Behaviour Scale for Sport

Physical training and planning								
The coach(es) most responsible for my physical training and conditioning...								
		never						**always**
1)	Provides me with a physical condition program in which I am confident	○	○	○	○	○	○	○
2)	Provides me with a physically challenging conditioning program	○	○	○	○	○	○	○
3)	Provides me with a detailed physical conditioning program	○	○	○	○	○	○	○
4)	Provides me with a plan for my physical preparation	○	○	○	○	○	○	○
5)	Ensures that training facilities and equipment are organized	○	○	○	○	○	○	○
6)	Provide me with structured training sessions	○	○	○	○	○	○	○
7)	Provides me with an annual training program	○	○	○	○	○	○	○
Technical skills								
The coach(es) most responsible for my technical skills...								
		never						**always**
8)	Provides me with advice while I'm performing a skill	○	○	○	○	○	○	○
9)	Gives me specific feedback for correcting technical errors	○	○	○	○	○	○	○
10)	Gives me reinforcement about correct technique	○	○	○	○	○	○	○
11)	Provides me with feedback that helps me improve my technique	○	○	○	○	○	○	○
12)	Provides visual examples to show how a skill should be done	○	○	○	○	○	○	○
13)	Uses verbal examples that describe how a skill should be done	○	○	○	○	○	○	○
14)	Makes sure I understand the techniques and strategies I'm being taught	○	○	○	○	○	○	○
15)	Provides me with immediate feedback	○	○	○	○	○	○	○
Mental preparation								
The coach(es) most responsible for my mental preparation...								
		never						**always**
16)	Provides advice on how to perform under pressure	○	○	○	○	○	○	○
17)	Provides advice on how to be mentally tough	○	○	○	○	○	○	○
18)	Provides advice on how to stay confident about my abilities	○	○	○	○	○	○	○
19)	Provides advice on how to stay positive about myself	○	○	○	○	○	○	○
20)	Provides advice on how to stay focused	○	○	○	○	○	○	○

Continute on the next page.

Source:[22]

Goal settings								
The coach(es) most responsible for my goal settings...								
		never						**always**
21)	Helps me identify strategies to achieve my goals	○	○	○	○	○	○	○
22)	Monitors my progress toward my goals	○	○	○	○	○	○	○
23)	Helps me set-short term goals	○	○	○	○	○	○	○
24)	Helps me identify target dates for attaining my goals	○	○	○	○	○	○	○
25)	Helps me set long-term goals	○	○	○	○	○	○	○
26)	Provides support to attain my goals	○	○	○	○	○	○	○
Competition strategies								
The coach(es) most responsible for my competition strategies...								
		never						**always**
27)	Helps me focus on the process of performing well	○	○	○	○	○	○	○
28)	Prepares me to face a variety of situations in competition	○	○	○	○	○	○	○
29)	Keeps me focused in competition	○	○	○	○	○	○	○
30)	Has a consistent routine at competition	○	○	○	○	○	○	○
31)	Deals with problems I may experience at competitions	○	○	○	○	○	○	○
32)	Shows confidence in my ability during competitions	○	○	○	○	○	○	○
33)	Ensures that facilities and equipment are organized for competition	○	○	○	○	○	○	○
Personal rapport								
The coach(es) most responsible for my...								
		never						**always**
34)	Show understanding for me as a person	○	○	○	○	○	○	○
35)	Is a good listener	○	○	○	○	○	○	○
36)	Is easily approachable about personal problem I might have	○	○	○	○	○	○	○
37)	Demonstrates concern for my whole self (i.e., other parts of my life than sport)	○	○	○	○	○	○	○
38)	Is trustworthy with my personal problems	○	○	○	○	○	○	○
39)	Maintains confidentiality regarding my personal life	○	○	○	○	○	○	○
Negative personal rapport								
My head coach...								
		never						**always**
40)	Uses fear in his/her coaching methods	○	○	○	○	○	○	○
41)	Yells at me when angry	○	○	○	○	○	○	○
42)	Disregards my opinion	○	○	○	○	○	○	○
43)	Shows favouritism towards others	○	○	○	○	○	○	○
44)	Intimidates me physically	○	○	○	○	○	○	○
45)	Uses power to manipulate me	○	○	○	○	○	○	○
46)	Makes personal comments to me that I find upsetting	○	○	○	○	○	○	○
47)	Spends more time coaching the best athletes	○	○	○	○	○	○	○

A.7.2 DirectLife Coaching

		nooit						**altijd**
1)	Mijn coach geeft advies over hoe ik onder druk of in moeilijke situaties kan presteren.	○	○	○	○	○	○	○
2)	Mijn coach geeft advies over hoe ik kan doorzetten.	○	○	○	○	○	○	○
3)	Mijn coach geeft advies over hoe ik vertrouwen kan blijven houden in mijn vaardigheden.:	○	○	○	○	○	○	○
4)	Mijn coach geeft advies over hoe ik positief over mezelf kan blijven.	○	○	○	○	○	○	○
5)	Mijn coach geeft advies over hoe ik gericht kan blijven (op mijn activiteitsdoelen).	○	○	○	○	○	○	○
6)	Mijn coach reikt me strategieën aan om mijn doelen te halen.	○	○	○	○	○	○	○
7)	Mijn coach houdt mijn voortgang richting mijn doel in de gaten.	○	○	○	○	○	○	○
8)	Mijn coach Helpt me om kortetermijn-doelen te stellen en actieplannen te maken.	○	○	○	○	○	○	○
9)	Mijn coach helpt me om richtdatums te kiezen om (tussen)doelen te behalen	○	○	○	○	○	○	○
10)	Mijn coach geeft me ondersteuning om mijn doelen te behalen.	○	○	○	○	○	○	○
11)	Mijn coach laat zien dat hij/zij me begrijpt.	○	○	○	○	○	○	○
12)	Mijn coach kan goed luisteren.	○	○	○	○	○	○	○
13)	Mijn coach is gemakkelijk benaderbaar over eventuele persoonlijke	○	○	○	○	○	○	○
14)	Mijn coach laat zien dat hij/zij om me geeft (dus ook andere aspecten van mijn leven en niet alleen lichamelijke activiteit)	○	○	○	○	○	○	○
15)	Mijn coach kan ik mijn persoonlijke problemen toevertrouwen.	○	○	○	○	○	○	○
16)	Mijn coach gaat vertrouwelijk om met mijn persoonlijke problemen.	○	○	○	○	○	○	○
17)	Mijn coach helpt me om mijn activiteitsprestaties te herkennen en vieren.	○	○	○	○	○	○	○
18)	Mijn coach helpt me om motivatie en inspiratie uit anderen te halen	○	○	○	○	○	○	○
19)	Mijn coach helpt me te ontdekken welke dingen het best voor mij werken om actiever te worden.	○	○	○	○	○	○	○
20)	Mijn coach heeft de juiste kennis en vaardigheden om goede coaching te geven.	○	○	○	○	○	○	○
21)	Mijn coach geeft coaching van goede kwaliteit.	○	○	○	○	○	○	○

Source: Philips

B Diaries

B.1 Day 1

Hoe actief bent u vandaag geweest?

1. Welke activiteiten op het gebied van lichaamsbeweging (wandelen, hond uitlaten, fietsen, trap lopen, sporten, etc.) heeft u vandaag gedaan, en wanneer was dat? Geef hieronder op de tijdslijn aan welke fysieke beweging en/of activiteiten u vandaag heeft gedaan, en welk tijdstip dit was.

2. Stickers plakken!

Wat zou een goed moment zijn om een herinnering te krijgen dat u meer zou kunnen bewegen? Plak op de tijdslijn een sticker met het alarm icoontje.

Wat zou een goed moment zijn om een bericht te krijgen dat aangeeft hoe actief u vandaag al geweest bent? Plak op de tijdslijn een sticker met het feedback icoontje.

3. Beschrijf hieronder wanneer u het liefst een activiteitsherinnering zou willen ontvangen en wanneer u het liefst een overzicht (van uw activiteitsniveau) zou willen zien. Dit kan bijvoorbeeld zijn; thuis, onderweg, op uw werk, tijdens het boodschappen doen, etc.

Een activiteitsherinnering ontvang ik het liefs wanneer ik:

_____

_____

Een activiteitsoverzicht bekijk ik het liefst wanneer ik:

_____

_____

4. Beschrijf hieronder wanneer zou u absoluut geen activiteitsherinnering willen ontvangen? Wanneer zou u absoluut geen activiteitsoverzicht willen ontvangen?

_____

_____

5. Kruis hieronder aan welk(e) bericht(en) u vandaag heeft ontvangen en omcirkel op welk apparaat u dit bericht heeft gelezen.

Bericht	**Apparaat**	**Opmerkingen**
		Wat vond u van het tijdstip? Was uw ervaring positief of negatief?
☐ Medicijnen vergeten		_____
☐ 100% medicijn inname		_____
☐ Activiteitsdoel gehaald		_____
☐ Aansluiten AM		_____
☐ Niet voldoende bewogen		_____

6. Kruis hieronder aan welk(e) overzicht(en) u vandaag heeft bekeken op welke apparaat. Beschrijf ook hoe vaak en wanneer u dit heeft gedaan.

Overzicht	**Apparaat**	**Hoe vaak en wanneer?**
☐ Medicijnoverzicht		_____
☐ Activiteitsoverzicht		_____
☐ Medicijnoverzicht		_____
☐ Activiteitsoverzicht		_____

7. Heeft u vandaag (technische) problemen ondervonden met het systeem?

_____

_____

8. Heeft u verder nog opmerkingen? Schrijf ze dan hieronder op.

_____

_____

B.2 Day 2

Wat doet u om uw medicijnen niet te vergeten?

1. Hoe zorgt u ervoor dat u uw medicatie op tijd neemt? Doet u dit met behulp van een post-it? Ligt uw medicatie op een vaste plek? Of heeft u een andere manier?

_____

_____

2. Vergelijk uw manier met het systeem dat u deze week gebruikt. Zou u het verschil in ervaring kunnen omschrijven en toelichten waarom u dit vindt?

_____

_____

3. Vindt u de berichten die u tot nu toe van het systeem heeft gekregen over uw medicijngebruik prettig of niet? Kunt u ook omschrijven waarom? (Als u geen berichten heeft gekregen over uw medicijn gebruik, geef dat dan hieronder aan)

_____

_____

4. Kruis hieronder aan welk(e) bericht(en) u vandaag heeft ontvangen en omcirkel op welk apparaat u dit bericht heeft gelezen.

Bericht	**Apparaat**	**Opmerkingen**
		Wat vond u van het tijdstip? Was uw ervaring positief of negatief?
☐ Medicijnen vergeten		_____
☐ 100% medicijn inname		_____
☐ Activiteitsdoel gehaald		_____
☐ Aansluiten AM		_____
☐ Niet voldoende bewogen		_____

5. Kruis hieronder aan welk(e) overzicht(en) u vandaag heeft bekeken op welke apparaat. Beschrijf ook hoe vaak en wanneer u dit heeft gedaan.

Overzicht	**Apparaat**	**Hoe vaak en wanneer?**
☐ Medicijnoverzicht		_____
☐ Activiteitsoverzicht		_____
☐ Medicijnoverzicht		_____
☐ Activiteitsoverzicht		_____

6. Heeft u vandaag (technische) problemen ondervonden met het systeem?

_____

_____

7. Heeft u verder nog opmerkingen? Schrijf ze dan hieronder op.

_____

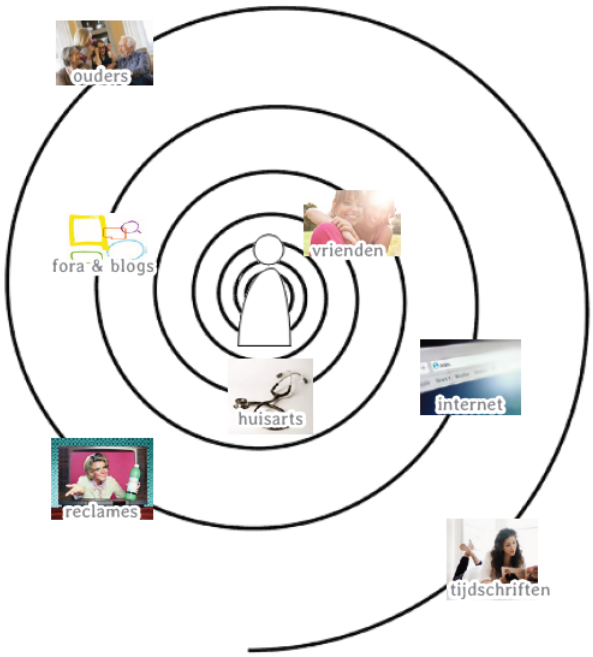
_____

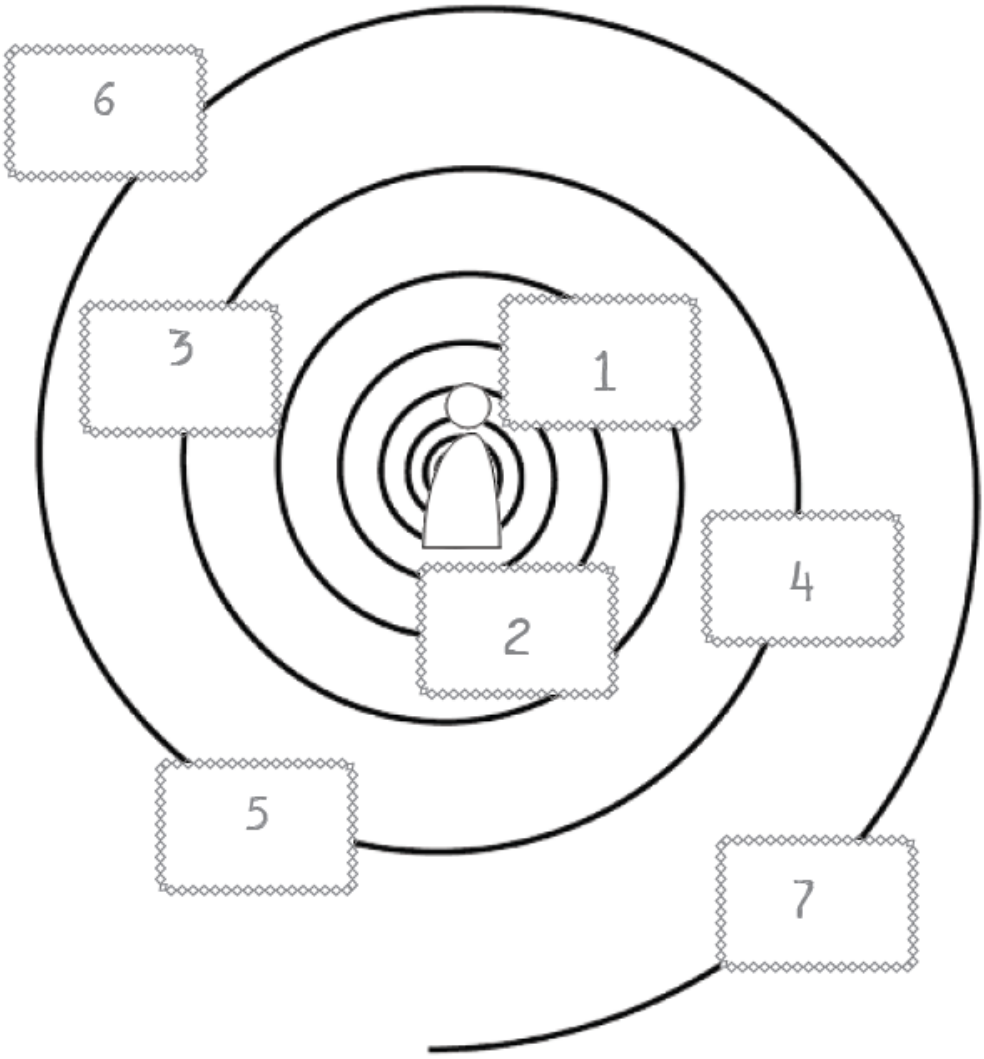
B.3 Day 3

Waar haalt u informatie over een gezonde leefstijl vandaan?

1. Vandaag gaat het over verschillende bronnen met informatie over een gezonde leefstijl en gezondheid die u mogelijk raadpleegt. Denk hierbij bijvoorbeeld aan de huisarts, vrienden, internet fora, tijdschriften, etc. Op stickervel 3 vindt u stickers voor elk van deze informatiebronnen.

In het midden van de cirkel hiernaast staat een poppetje afgebeeld. Dit bent u. Gebruik nu de stickertjes en plak deze op de genummerde vakjes in de spiraal. Hoe dichter u de sticker in het midden plakt, hoe meer waarde u hecht aan deze informatie bron. Dus de bron die u het belangrijkst vindt, komt op vakje 1, en de bron die voor u het minst belangrijk is, kunt u op vakje met nummer 7 plakken. U vindt hieronder een voorbeeld.





2. Kunt u voor elke bron aangeven wat dit voor u betekent en wat u van deze informatie vindt.

Huisarts _____

Vrienden _____

Ouders _____

Internet _____

Internet fora/blogs _____

Tijdschriften _____

Reclames _____

3. Kruis hieronder aan welk(e) bericht(en) u vandaag heeft ontvangen en omcirkel op welk apparaat u dit bericht heeft gelezen.

Bericht	**Apparaat**	**Opmerkingen**
		Wat vond u van het tijdstip? Was uw ervaring positief of negatief?
☐ Medicijnen vergeten		
☐ 100% medicijn inname		
☐ Activiteitsdoel gehaald		
☐ Aansluiten AM		
☐ Niet voldoende bewogen		

4. Kruis hieronder aan welk(e) overzicht(en) u vandaag heeft bekeken op welke apparaat. Beschrijf ook hoe vaak en wanneer u dit heeft gedaan.

Overzicht	**Apparaat**	**Hoe vaak en wanneer?**
☐ Medicijnoverzicht		
☐ Activiteitsoverzicht		
☐ Medicijnoverzicht		
☐ Activiteitsoverzicht		

5. Heeft u vandaag (technische) problemen ondervonden met het systeem?

6. Heeft u verder nog opmerkingen? Schrijf ze dan hieronder op.

B.4 Day 4

Wanneer neemt u uw medicijnen in?

1. Welke activiteiten op het gebied van lichaamsbeweging (wandelen, hond uitlaten, fietsen, trap lopen, sporten, etc.) heeft u vandaag gedaan, en wanneer was dat? Geef hieronder op de tijdslijn aan welke fysieke beweging en/of activiteiten u vandaag heeft gedaan, en welk tijdstip dit was.

2. Stickers plakken!

Wat zou een goed moment zijn om een herinnering te krijgen dat u meer zou kunnen bewegen? Plak op de tijdslijn een sticker met het alarm icoontje.

Wat zou een goed moment zijn om een bericht te krijgen dat aangeeft hoe actief u vandaag al geweest bent? Plak op de tijdslijn een sticker met het feedback icoontje.

3. Beschrijf tijdens welke bezigheden u een medicijn herinnering zou willen krijgen (thuis, onderweg, op uw werk, tijdens het doen van de boodschappen, etc.).

4. Tijdens welke bezigheden zou u absoluut geen medicijn herinnering willen ontvangen?

5. Medicijn herinneringen worden momenteel een half uur na het moment dat u heeft ingesteld verstuurd. Wat vindt u hiervan? *Als u uw medicijnen bijvoorbeeld om 10u inneemt, zult u om half 11 een herinnering ontvangen van ons systeem als u dat dan nog niet gedaan heeft.*

6. Hoe lang zou u van te voren en/of achteraf een medicijn herinnering willen krijgen?

7. Kruis hieronder aan welk(e) bericht(en) u vandaag heeft ontvangen en omcirkel op welk apparaat u dit bericht heeft gelezen.

Bericht	**Apparaat**	**Opmerkingen**
		Wat vond u van het tijdstip? Was uw ervaring positief of negatief?
☐ Medicijnen vergeten		
☐ 100% medicijn inname		
☐ Activiteitsdoel gehaald		
☐ Aansluiten AM		
☐ Niet voldoende bewogen		

8. Kruis hieronder aan welk(e) overzicht(en) u vandaag heeft bekeken op welke apparaat. Beschrijf ook hoe vaak en wanneer u dit heeft gedaan.

Overzicht	**Apparaat**	**Hoe vaak en wanneer?**
☐ Medicijnoverzicht		
☐ Activiteitsoverzicht		
☐ Medicijnoverzicht		
☐ Activiteitsoverzicht		

9. Heeft u vandaag (technische) problemen ondervonden met het systeem?

10. Heeft u verder nog opmerkingen? Schrijf ze dan hieronder op.

B.5 Day 5

Vandaag gaat het over de verschillende functies van het systeem

1. Welke onderdelen en functies van het systeem heeft u vandaag gebruikt? Hieronder kunt u aankruisen welke onderdelen van het systeeem u heeft gebruikt.

Onderdelen van het systeem:

☐ Medicijndoosje

☐ Activiteitsmeter

Hieronder kunt u aankruisen welke functies u heeft gebruikt op de computer en de telefoon.

Functies van het systeem:

Op computer	Op telefoon
☐ Berichten ontvangen	☐ Berichten ontvangen
☐ Overzicht van activiteit	☐ Overzicht van activiteit
☐ Overzicht van medicijninname	☐ Overzicht van medicijninname

2. Geef op de plaatjes op de volgende bladzijdes aan welke onderdelen van het systeem u prettig vond, en welke minder. U kunt hiervoor de bijgevoegde smiley stickers gebruiken. Kunt u ook aangeven wat hiervan de reden is? U kunt uw opmerkingen er gewoon bijschrijven.

Table B.1: Smileys to choose from

3. Kruis hieronder aan welk(e) bericht(en) u vandaag heeft ontvangen en omcirkel op welk apparaat u dit bericht heeft gelezen.

Bericht	**Apparaat**	**Opmerkingen**
		Wat vond u van het tijdstip? Was uw ervaring positief of negatief?
☐ Medicijnen vergeten		_____
☐ 100% medicijn inname		_____
☐ Activiteitsdoel gehaald		_____
☐ Aansluiten AM		_____
☐ Niet voldoende bewogen		_____

4. Kruis hieronder aan welk(e) overzicht(en) u vandaag heeft bekeken op welke apparaat. Beschrijf ook hoe vaak en wanneer u dit heeft gedaan.

Overzicht	**Apparaat**	**Hoe vaak en wanneer?**
☐ Medicijnoverzicht		_____
☐ Activiteitsoverzicht		_____
☐ Medicijnoverzicht		_____
☐ Activiteitsoverzicht		_____

5. Heeft u vandaag (technische) problemen ondervonden met het systeem?

_____

_____

6. Heeft u verder nog opmerkingen? Schrijf ze dan hieronder op.

_____

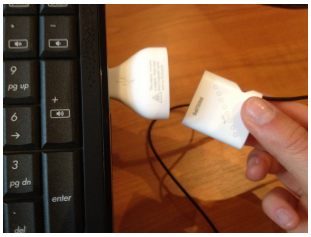
_____

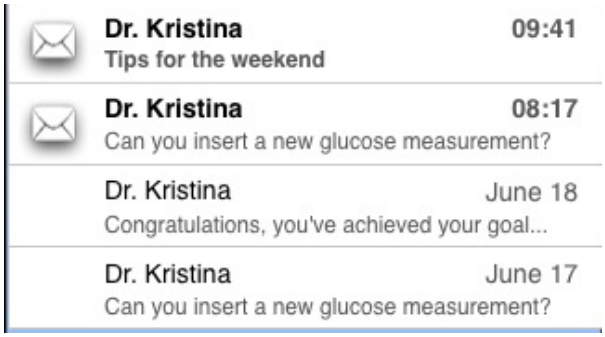
B.6 Day 6

Vandaag kunt u aangeven welke woorden u bij het systeem vindt passen

1. Op het stickervel van dag 6 staat een groot aantal woorden die eigenschappen beschrijven. Plak de woorden op bij de verschillende onderdelen van het systeem. De onderdelen staan op de deze pagina en de volgende twee pagina's.

Voorbeeld: wanneer u de smartphone applicatie handig vindt, dan plakt u de sticker waar 'handig' opstaat, op de afbeelding van de smartphone hiernaast.





Menselijk	Goedkoop	Vergroot de afstand tot anderen	Uitdagend	Onhandelbaar
Samenhangend	GEÏNTEGREERD	stijlvol	Motiverend	Nieuw
Onprofessioneel	Lelijk	Creatief	Aantrekkelijk	ONAANGENAAM
Handig	Ingewikkeld	Duidelijk gestructureerd	Brutaal	Smakeloos
Onvoorspelbaar	Onhandig	Prettig	Toonbaar	saai
Innovatief	AANGENAAM	Onprofessioneel	Afstandelijk	VERWARREND
Ontmoedigend	Onaangenaam	Fantasieloos	Uitnodigend	Voorspelbaar
Vervreemd	MOEILIJK TE HANTEREN	Alledaags	Op zichzelf staand	EENVOUDIG TE HANTEREN

Figure B.1: All words to choose from

2. De woorden hieronder zijn tegengesteld aan elkaar. Kunt u per regel aankruisen wat u het meest bij het systeem vindt passen?

menselijk	○	○	○	○	○	○	○	technisch
isolerend	○	○	○	○	○	○	○	verbindend
plezierig	○	○	○	○	○	○	○	onplezierig
inventief	○	○	○	○	○	○	○	traditioneel
simpel	○	○	○	○	○	○	○	ingewikkeld
professioneel	○	○	○	○	○	○	○	onprofessioneel
lelijk	○	○	○	○	○	○	○	aantrekkelijk
practisch	○	○	○	○	○	○	○	onpractisch
aardig	○	○	○	○	○	○	○	vervelend
omslachtig	○	○	○	○	○	○	○	eenvoudig
stijlvol	○	○	○	○	○	○	○	smaakloos
voorspelbaar	○	○	○	○	○	○	○	onvoorspelbaar
goedkoop	○	○	○	○	○	○	○	chic
vervreemdend	○	○	○	○	○	○	○	integrerend
brengt mij dichter bij mensen	○	○	○	○	○	○	○	isoleerd mij van mensen
onaantrekkelijk	○	○	○	○	○	○	○	toonbaar
afwijzend	○	○	○	○	○	○	○	verwelkomend
fantasieloos	○	○	○	○	○	○	○	creatief
goed	○	○	○	○	○	○	○	slecht
verwarrend	○	○	○	○	○	○	○	goed gestructureerd
afstotend	○	○	○	○	○	○	○	aantrekkelijk
gewaagd	○	○	○	○	○	○	○	voorzichtig
vernieuwend	○	○	○	○	○	○	○	conservatief
saai	○	○	○	○	○	○	○	boeiend
gemakkelijk	○	○	○	○	○	○	○	uitdagend
motiverend	○	○	○	○	○	○	○	ontmoedigend
nieuw	○	○	○	○	○	○	○	allerdaags
onhandelbaar	○	○	○	○	○	○	○	handelbaar

3. Kruis hieronder aan welk(e) bericht(en) u vandaag heeft ontvangen en omcirkel op welk apparaat u dit bericht heeft gelezen.

Bericht	**Apparaat**	**Opmerkingen**
		Wat vond u van het tijdstip? Was uw ervaring positief of negatief?
☐ Medicijnen vergeten		_____
☐ 100% medicijn inname		_____
☐ Activiteitsdoel gehaald		_____
☐ Aansluiten AM		_____
☐ Niet voldoende bewogen		_____

4. Kruis hieronder aan welk(e) overzicht(en) u vandaag heeft bekeken op welke apparaat. Beschrijf ook hoe vaak en wanneer u dit heeft gedaan.

Overzicht	**Apparaat**	**Hoe vaak en wanneer?**
☐ Medicijnoverzicht		_____
☐ Activiteitsoverzicht		_____
☐ Medicijnoverzicht		_____
☐ Activiteitsoverzicht		_____

5. Heeft u vandaag (technische) problemen ondervonden met het systeem?

_____

_____

6. Heeft u verder nog opmerkingen? Schrijf ze dan hieronder op.

_____

_____

B.7 Day 7

Vandaag is de laatste dag, we zouden u willen vragen om het gehele systeem te beoordelen

1. Zou u de vragen hieronder willen invullen? Wees alstublieft eerlijk in uw antwoorden! U krijgt de mogelijkheid uw antwoorden toe te lichten tijdens het afsluitende interview.

	helemaal mee eens				**helemaal mee oneens**
Ik zou dit system graag met regelmaat willen gebruiken.	○	○	○	○	○
Ik vind het systeem onnodig ingewikkeld.	○	○	○	○	○
Ik vind het systeem gemakkelijk om te gebruiken.	○	○	○	○	○
Ik heb de hulp van een technisch persoon nodig om dit systeem te kunnen gebruiken.	○	○	○	○	○
Ik vind dat de verschillende functies van het systeem goed geïntegreerd waren.	○	○	○	○	○
Ik vind het systeem niet consistent.	○	○	○	○	○
Ik denk dat mensen gemakkelijk kunnen leren hoe ze dit systeem kunnen gebruiken.	○	○	○	○	○
Ik vind het systeem erg moeizaam in gebruik.	○	○	○	○	○
Ik voel me erg zelfverzekerd wanneer ik het systeem gebruik.	○	○	○	○	○
Ik heb een boel dingen moeten leren voordat ik aan de slag kon met het systeem.	○	○	○	○	○

2. Kruis hieronder aan welk(e) bericht(en) u vandaag heeft ontvangen en omcirkel op welk apparaat u dit bericht heeft gelezen.

Bericht	**Apparaat**	**Opmerkingen**
		Wat vond u van het tijdstip?
		Was uw ervaring positief of negatief?
☐ Medicijnen vergeten		_____
☐ 100% medicijn inname		_____
☐ Activiteitsdoel gehaald		_____
☐ Aansluiten AM		_____
☐ Niet voldoende bewogen		_____

3. Kruis hieronder aan welk(e) overzicht(en) u vandaag heeft bekeken op welke apparaat. Beschrijf ook hoe vaak en wanneer u dit heeft gedaan.

Overzicht	**Apparaat**	**Hoe vaak en wanneer?**
☐ Medicijnoverzicht		_____
☐ Activiteitsoverzicht		_____
☐ Medicijnoverzicht		_____
☐ Activiteitsoverzicht		_____

4. Heeft u vandaag (technische) problemen ondervonden met het systeem?

_____

_____

5. Heeft u verder nog opmerkingen? Schrijf ze dan hieronder op.

_____

_____

C Evaluation questionnaires

C.1 Welcome and introduction questions

Nummer van je bewegingsmeter:

_____

Leeftijd:

_____

Geslacht:

- ○ Vrouw
- ○ Man

Waar gebruik je je smartphone vooral voor?

_____

_____

_____

Hoeveel uur per dag denk je dat je zit tijdens je werk/studie?
Maak hier een schatting in hele uren.

_____

Vind je dat je normaal gesproken genoeg beweegt?

- ○ Ja, ik beweeg genoeg
- ○ Nee, ik beweeg niet genoeg
- ○ Anders: _____

Ben je gemiddeld genomen elke dag 30 minuten redelijk actief?

- ○ Ik ben elke dag meer dan 30 minuten redelijk actief
- ○ Ik ben elke dag net 30 minuten redelijk actief
- ○ Ik ben elke dag minder dan 30 minuten redelijk actief

Beoefen je een sport?
Zo ja, wat doe je en hoe vaak?

_____

_____

Kun je aangeven wat redenen voor je zouden zijn op dit moment om je bewegingspatroon aan te passen?
(meerdere antwoorden zijn mogelijk)

- ☐ Stressreductie
- ☐ Energie verhogen
- ☐ Voor de lol
- ☐ Een uitdaging zoeken
- ☐ Bewonderd worden
- ☐ Sociale bezigheid
- ☐ Winkansen verhogen voor sporten
- ☐ Op doktersadvies
- ☐ Ziektepreventie
- ☐ Algemene gezondheid verbeteren
- ☐ Gewichtscontrole
- ☐ Er beter uit zien
- ☐ Fitheid verhogen
- ☐ Leniger/flexibeler worden
- ☐ Anders: _____

Kun je aangeven wat je ervaring is?
Op een schaal van 1 tot 10 (1 = helemaal geen, 10 = heel veel)

	helemaal geen									**heel veel**
	1	2	3	4	5	6	7	8	9	10
Smartphones	○	○	○	○	○	○	○	○	○	○
Systemen die bedoeld zijn om je gedrag aan te passen	○	○	○	○	○	○	○	○	○	○
Systemen die je coachen	○	○	○	○	○	○	○	○	○	○
Embodied converstational agent (virtuele karakters)	○	○	○	○	○	○	○	○	○	○
Mobiel internet	○	○	○	○	○	○	○	○	○	○

Verwachtingen

De volgende vragen gaan over je verwachtingen over het systeem.

Kun je aangeven wat je van het systeem verwacht? Denk hierbij aan; Wat zal het systeem doen? Wat zal ik er aan hebben? Ga ik het nuttig vinden? etc. etc.?

_____

_____

_____

_____

_____

_____

Kun je aangeven wat je verwachtingen zijn van het system met behulp van de volgende stellingen.
Op een schaal van 1 tot 7 (1 = helemaal mee oneens, 7 = helemaal mee eens)

	helemaal mee oneens						**helemaal mee eens**
Het systeem maakt mijn leven interessanter	○	○	○	○	○	○	○
Ik vind het fijn om met het systeem te werken	○	○	○	○	○	○	○
Ik vind het een goed idee om het systeem te gebruiken	○	○	○	○	○	○	○
Werken met het systeem is leuk	○	○	○	○	○	○	○
Ik vind het systeem intimiderend	○	○	○	○	○	○	○
Ik ben bang om het systeem te gebruiken	○	○	○	○	○	○	○
Met het systeem beweeg ik meer	○	○	○	○	○	○	○
Ik vind het systeem handig om meer te bewegen	○	○	○	○	○	○	○
Ik volg het advies van het systeem op	○	○	○	○	○	○	○
Met het systeem bereik ik sneller mijn doel om meer te bewegen	○	○	○	○	○	○	○
Ik vertrouw het advies wat het systeem mij geeft	○	○	○	○	○	○	○
Als ik het systeem gebruik, ben ik bang fouten te maken of iets kapot te maken	○	○	○	○	○	○	○

C.2 Timing of messages

Nummer van je bewegingsmeter: _____

De volgende vragen gaan over de timing van de berichten die je hebt gekregen. Probeer zo duidelijk mogelijk op te schrijven waarom je iets vindt.

Wanneer heb je vandaag een bericht ontvangen?
Was dit om 12:00 uur, rond het avondeten, of andere momenten?

Wat vond je van het tijdstip waarop je de afgelopen dagen berichten hebt ontvangen?
Vond je deze timing goed, of zou het volgens jou anders moeten?

Wanneer op de dag zou je een herinnering willen ontvangen om actiever te zijn?

Wanneer op de dag zou je GEEN herinnering willen ontvangen om actiever te zijn?

Wanneer bekijk je je het overzicht van je activiteit?

Het systeem algemeen

De volgende vragen gaan over het hele systeem.

Vul in in hoeverre je het met de volgende stellingen eens bent.
1 = heel erg niet eens, 5 = heel erg eens

	helemaal mee oneens				**helemaal mee eens**
	1	2	3	4	5
Ik vertrouw het advies wat het systeem mij geeft.	○	○	○	○	○
Ik denk dat het systeem mij helpt wanneer ik dit nodig vind.	○	○	○	○	○
Ik denk dat het systeem alleen doet wat ik op dat moment nodig heb.	○	○	○	○	○
Ik volg het advies van het systeem op.	○	○	○	○	○
Ik denk dat het systeem zich aan kan passen aan wat ik nodig heb .	○	○	○	○	○

Vul in in hoeverre je het met de volgende stellingen eens bent.
1 = heel erg niet eens, 7 = heel erg eens

	helemaal mee oneens						**helemaal mee eens**
	1	2	3	4	5	6	7
Als ik het systeem gebruik, ben ik bang fouten te maken of iets kapot te maken	○	○	○	○	○	○	○
Het systeem maakt mijn leven interessanter	○	○	○	○	○	○	○
Ik vind het een goed idee om het systeem te gebruiken	○	○	○	○	○	○	○
Ik vind het systeem intimiderend	○	○	○	○	○	○	○
Werken met het systeem is leuk	○	○	○	○	○	○	○
Ik ben bang om het systeem te gebruiken	○	○	○	○	○	○	○
Ik vind het fijn om met het systeem te werken	○	○	○	○	○	○	○

C.3 Content of the messages

Nummer van je bewegingsmeter: _____

De volgende vragen gaan over de inhoud van de berichten die je de afgelopen dagen hebt gekregen. Probeer zo duidelijk mogelijk op te schrijven waarom je iets vindt.

Wat vind je van het aantal berichten?
Waren dit er genoeg, te weinig, of juiste teveel, en waarom vind je dit?

Wat vindt je van de inhoud van de berichten?
Geef toelichting waarom je dit vindt.

Welke berichten vond je prettig om te ontvangen?
Schrijf de berichten op en geef toelichting waarom je dit vindt.

Welke berichten vond je ONprettig om te ontvangen?
Schrijf de berichten op en geef toelichting waarom je dit vindt.

Was volgens je de inhoud van de berichten juist?

Vond je de berichten motiverend?

Wat voor soort berichten zou je nog meer willen ontvangen?
Geef eventueel hierbij een voorbeeld.

Het systeem algemeen

De volgende vragen gaan over het hele systeem.

Vul in in hoeverre je het met de volgende stellingen eens bent.
1 = heel erg niet eens, 7 = heel erg eens

	helemaal mee oneens						**helemaal mee eens**
	1	2	3	4	5	6	7
Met het systeem beweeg ik meer.	○	○	○	○	○	○	○
Ik heb alles wat ik nodig heb om het systeem goed te gebruiken.	○	○	○	○	○	○	○
Ik weet genoeg van het systeem om hier goed gebruik van te maken.	○	○	○	○	○	○	○
Ik vond het systeem flexibel om interactie mee te hebben.	○	○	○	○	○	○	○
Het systeem maakt het gemakkelijker voor mij om meer te bewegen.	○	○	○	○	○	○	○
Met het systeem bereik ik sneller mijn doel om meer te bewegen.	○	○	○	○	○	○	○
Ik zou het systeem graag willen blijven gebruiken.	○	○	○	○	○	○	○

Vul in in hoeverre je het met de volgende stellingen eens bent.
1 = heel erg niet eens, 7 = heel erg eens

	helemaal mee oneens						**helemaal mee eens**
	1	2	3	4	5	6	7
Ik vind het systeem handig voor mij.	○	○	○	○	○	○	○
Ik vond het gemakkelijk om het systeem te laten doen wat ik wilde.	○	○	○	○	○	○	○
Ik vind het systeem handig om meer te bewegen.	○	○	○	○	○	○	○
Ik vond het gemakkelijk om het systeem te gebruiken.	○	○	○	○	○	○	○
Mijn interactie met het systeem was begrijpelijk en duidelijk.	○	○	○	○	○	○	○
Het is gemakkelijk voor mij om het systeem te leren gebruiken.	○	○	○	○	○	○	○
Ik zou het handig vinden om het systeem te hebben.	○	○	○	○	○	○	○

C.4 Coaching

De volgende vragen gaan over de coaching die je hebt gekregen.

Nummer van je bewegingsmeter: _____

Hieronder staan enkele stellingen over de coaching. Geef aan in hoeverre je het eens bent met de stelling.
1 = nooit, 5 = altijd

	nooit				**altijd**
	1	2	3	4	5
Mijn coach geeft coaching van goede kwaliteit.	○	○	○	○	○
Mijn coach geeft advies over hoe ik gericht kan blijven (op mijn activiteitsdoelen).	○	○	○	○	○
Mijn coach geeft advies over hoe ik kan doorzetten.	○	○	○	○	○
Mijn coach helpt me om korte termijn doelen te stellen en actieplannen te maken.	○	○	○	○	○
Mijn coach heeft de juiste kennis en vaardigheden om goede coaching te geven.	○	○	○	○	○
Mijn coach helpt me om richtdatums te kiezen om (tussen)doelen te behalen.	○	○	○	○	○
Mijn coach reikt me strategieën aan om mijn doelen te halen.	○	○	○	○	○

Hieronder staan enkele stellingen over de coaching. Geef aan in hoeverre je het eens bent met de stelling.
1 = nooit, 5 = altijd

	nooit				**altijd**
	1	2	3	4	5
Mijn coach geeft advies over hoe ik vertrouwen kan blijven houden in mijn vaardigheden.	○	○	○	○	○
Mijn coach helpt me om mijn activiteitsprestaties te herkennen en vieren.	○	○	○	○	○
Mijn coach helpt me te ontdekken welke dingen het best voor mij werken om actiever te worden.	○	○	○	○	○
Mijn coach houdt mijn voortgang richting mijn doel in de gaten.	○	○	○	○	○
Mijn coach helpt me om motivatie en inspiratie uit anderen te halen.	○	○	○	○	○
Mijn coach geeft advies over hoe ik positief over mezelf kan blijven.	○	○	○	○	○
Mijn coach geeft me ondersteuning om mijn doelen te halen.	○	○	○	○	○

Laat zien hoe je over de coaching denkt door de volgende vragen te beantwoorden.
Nummers 1 en 7 geven een heel sterk gevoel aan. Nummers 2 en 6 een redelijk sterk gevoel. Nummer 3 en 5 een zwak gevoel. Nummer 4 geeft een neutraal gevoel aan. Vul dit in en denk er niet te veel over na, er is geen goed of fout antwoord.

	1	2	3	4	5	6	7	
vervelend	○	○	○	○	○	○	○	aardig
onvriendelijk	○	○	○	○	○	○	○	vriendelijk
eerlijk	○	○	○	○	○	○	○	oneerlijk
onwetend	○	○	○	○	○	○	○	op de hoogte
intelligent	○	○	○	○	○	○	○	dom
leek	○	○	○	○	○	○	○	deskundig
betrouwbaar	○	○	○	○	○	○	○	onbetrouwbaar
egoïstisch	○	○	○	○	○	○	○	onbaatzuchtig
ongeschikt	○	○	○	○	○	○	○	geschikt
waardevol	○	○	○	○	○	○	○	waardeloos
plezierig	○	○	○	○	○	○	○	onplezierig
deugdzaam	○	○	○	○	○	○	○	zondig

C.5 End questions

Nummer van je bewegingsmeter: _____

Geef aan met behulp van de woordparen wat je de beste beschrijving voor het hele systeem vind.

	1	2	3	4	5	6	7	
menselijk	○	○	○	○	○	○	○	technisch
isolerend	○	○	○	○	○	○	○	verbindend
aangenaam	○	○	○	○	○	○	○	onaangenaam
origineel	○	○	○	○	○	○	○	conventioneel
eenvoudig	○	○	○	○	○	○	○	ingewikkeld
professioneel	○	○	○	○	○	○	○	amateuristisch
lelijk	○	○	○	○	○	○	○	mooi
praktisch	○	○	○	○	○	○	○	onpraktisch
sympathiek	○	○	○	○	○	○	○	onsympathiek
omslachtig	○	○	○	○	○	○	○	direct

Geef aan met behulp van de woordparen wat je de beste beschrijving voor het hele systeem vind.

	1	2	3	4	5	6	7	
stijlvol	○	○	○	○	○	○	○	stijlloos
voorspelbaar	○	○	○	○	○	○	○	onvoorspelbaar
goedkoop	○	○	○	○	○	○	○	waardevol
buitensluitend / vervreemdend	○	○	○	○	○	○	○	integrerend / inbegrepen
brengt mij dichter bij mensen	○	○	○	○	○	○	○	isoleert mij van mensen
niet toonbaar	○	○	○	○	○	○	○	toonbaar
afwijzend	○	○	○	○	○	○	○	uitnodigend
fantasieloos	○	○	○	○	○	○	○	creatief
goed	○	○	○	○	○	○	○	slecht

Geef aan met behulp van de woordparen wat je de beste beschrijving voor het hele systeem vindt.

	1	2	3	4	5	6	7	
verwarrend	○	○	○	○	○	○	○	overzichtelijk
afstotend	○	○	○	○	○	○	○	aantrekkelijk
moedig	○	○	○	○	○	○	○	voorzichtig
innovatief	○	○	○	○	○	○	○	conservatief
saai	○	○	○	○	○	○	○	boeiend
eenvoudig	○	○	○	○	○	○	○	uitdagend
motiverend	○	○	○	○	○	○	○	ontmoedigend
vernieuwend	○	○	○	○	○	○	○	alledaags
onhandelbaar	○	○	○	○	○	○	○	handelbaar

Geef hieronder aan in hoeverre je je kunt vinden in de volgende beweringen.
1 = heel erg niet eens, 5 = heel erg eens

	nooit				**altijd**
	1	2	3	4	5
Ik denk dat ik dit systeem regelmatig zou willen gebruiken	○	○	○	○	○
Ik vond het systeem onnodig ingewikkeld	○	○	○	○	○
Ik vond dat het systeem gemakkelijk te gebruiken was	○	○	○	○	○
Ik denk dat ik hulp nodig heb van een technisch persoon om het systeem te kunnen gebruiken	○	○	○	○	○
Ik vond dat de verschillende functies van het systeem goed geïntegreerd waren	○	○	○	○	○
Ik vond dat er te veel inconsistentie in het systeem was	○	○	○	○	○
Ik kan me voorstellen dat de meeste mensen gemakkelijk kunnen leren het systeem te gebruiken	○	○	○	○	○
Ik vond het systeem omslachtig om te gebruiken	○	○	○	○	○
Ik voelde me zelfverzekerd wanneer ik het systeem gebruikte	○	○	○	○	○
Ik moet veel dingen leren voordat ik het systeem kon gebruiken	○	○	○	○	○

D Interviews

D.1 Interview script SmarcoS-diabetic

Binnenkomst.

Kort small talk houden.

Zeggen dat we nu met het eind-interview beginnen en dit op nemen.

Zeggen dat we met wat algemene vragen beginnen:

- Wat vond u van het systeem? (algemeen)
- Heeft het systeem aan uw verwachtingen voldaan? (verwachting)
- Wat vond u prettig aan het systeem? (algemeen)
- Wat vond u vervelend of irritant aan het systeem? (algemeen)
- Bent u tevreden met de hoeveelheid tijd die nodig was om het systeem te gebruiken? (algemeen)
- Bent u tevreden over het gebruiksgemak van het systeem? (algemeen)
- Zou u een soort gelijke coaching zelf ook willen gebruiken? (algemeen)
- Denkt u dat dit soort coaching u zou helpen om een gezondere leefstijl aan te nemen? (algemeen)

Zeggen dat we nu iets dieper op de functies van het systeem in gaan:

- Wat vond je wel goed aan de functies van het systeem? (funcionaliteit)
  - Mogelijke functies (deze gelden natuurlijk ook voor de volgende vragen):
    - Berichten ontvangen
    - Overzicht van activiteit
    - Overzicht van medicijninname
- Wat vond je niet goed aan de functies van het systeem? (funcionaliteit)
- Welke functies heeft u het meest gebruikt(welke onderdelen)? (dagelijks gebruik/funcionaliteit)
  - Waarom deze functie?
  - Wanneer heeft u deze functie gebruikt?
- Welke functies heeft u het minst of helemaal niet gebruikt? (dagelijks gebruik/funcionaliteit)
  - Waarom deze functie?
  - Wanneer heeft u deze functie gebruikt? (als van toepassing)
- Kunt u voor elke functie aangeven op welk apparaat u deze heeft gebruikt? (dagelijks gebruik/ funcionaliteit)
- Is er een functie dat het systeem volgens u op dit moment mist? (funcionaliteit)
- Had u het gevoel controle te hebben over wat het systeem deed wanneer u het gebruikte? (dagelijks gebruik)
- Is er iets fout gegaan met het systeem of heeft u nog fouten ontdekt? (funcionaliteit)

Zeggen dat we nu gaan kijken naar de berichten:

- Welke berichten vond u prettig om te ontvangen? (coaching/feedback)
  - Doorvragen waarom?
- Welke berichten vond u niet prettig om te ontvangen? (coaching/feedback)
  - Doorvragen waarom?

- Wat vindt u van de timing van de berichten? (coaching/feedback)
  - Vraag door voor zowel feedback berichten als herinneringen, en waarom?
- Wat vindt u van de hoeveelheid berichten? (coaching/feedback)
  - Vraag door voor zowel feedback berichten als herinneringen, en waarom?
- Wat vindt u van de keuze van het apparaat? (coaching/feedback)
  - Vraag door voor zowel feedback berichten als herinneringen, en waarom?
- In welke situatie zou u graag een bericht willen krijgen? (coaching/feedback)
  - Vraag door voor zowel herinnering als feedback
- In welke situatie zou u niet graag een bericht willen krijgen? (coaching/feedback)
  - Vraag door voor zowel herinnering als feedback
- Vindt u het voldoende dat u berichten in tekstvorm krijgt, of zou u dit liever op een andere manier krijgen? (coaching/feedback)
  - Welke manier (voorbeeld, spraak, filmpje, etc.)?
- Heeft u het idee dat de berichten die u krijgt juist zijn? (coaching/feedback)

Zeggen dat we nu naar de verschillende apparaten gaan kijken:

- U gebruikt het systeem op twee verschillende apparaten, wat waren in uw ogen de belangrijkste verschillen? (inter-usability)
- Kon u dezelfde functies gemakkelijk terug vinden op de verschillende apparaten? (inter-usability)
- Heeft u het gevoel dat alle apparaten gesynchroniseerd waren (alle apparaten weten op hetzelfde moment de zelfde informatie)? (inter-usability)
- Heeft u het gevoel dat u met hetzelfde systeem communiceert, ondanks de twee verschillende apparten? (inter-usability)

Loop nu samen het dagboek door, vraag door op dingen die je opvallen, of waar de deelnemer over uitwijdt.
Zeggen dat nu de laatste vragen komen:

- Het systeem weet van alles. Wat denkt u dat het weet? (privacy)
  - Doorvragen wat de deelnemer hier van vindt of bezwaar heeft dat het systeem deze dingen weet . Begin met de dingen die de deelnemer zelf heeft aangegeven. Geef daarna aan wat het systeem nog meer weet en vraag wat de deelnemer hier van vindt.
  - Mogelijk antwoorden:
    * waar de deelnmer is
    * wat de deelnemer doet
    * hoeveel de deelnemer beweegt
    * of de deelnemer zijn medicijnen heeft genomen

Vraag nogmaals:

- Wat vond u van het systeem? (algemeen)(dit kan veranderd zijn omdat de deelnemer aan de hand van de vragen hier meer overna heeft kunnen denken)
- Heeft u nog suggesties of verbeteringen voor het systeem? (suggeties/verbeteringen)
- Heeft u verder nog iets gemist?

Dit wat het einde van het interview.

Bedanken de deelnemer voor de medewerking, vraag of ze op de hoogte willen worden gehouden.

D.2 Interview script SmarcoS-office worker

Verwachtingen

- Heeft het systeem aan uw verwachtingen voldaan? (gebaseerd op antwoorden van introductie gesprek)

Algemene indruk

- Zou u een soort gelijke coaching zelf ook willen gebruiken?
  - Als nee, wat moet er veranderd worden?
- Denkt u dat dit soort coaching u zou helpen om een gezondere leefstijl aan te nemen?
  - Als nee, wat moet er veranderd worden?
- Bent u tevreden met de hoeveelheid tijd die nodig was om het systeem te gebruiken?
- Bent u tevreden over het gebruiksgemak van het systeem?

Dagelijks gebruik

- Welke functionaliteiten heeft u het meest gebruikt? Waarom? Wanneer?
- Welke functies heeft u het minst of wellicht helemaal niet gebruikt? Waarom? Wanneer?
- Is er iets fout gaan met het systeem of heeft u nog fouten ontdekt?
- Wanneer u het systeem gebruikte, had u het gevoel controle te hebben over wat het systeem deed?
- Hebt u het gevoel dat het systeem u voldoende informatie geeft over wat er gaande is?
- Is er een functionaliteit dat het systeem op dit moment mist?

Coaching / Feedback

- In welke situaties zou u graag een herinnering of support willen krijgen?
- In welke situaties zou u absoluut niet graag een herinnering of support willen krijgen?
- Welke berichten (of herinneringen) vond u prettig om te ontvangen? Waarom?
- Welke berichten (of herinneringen) vond u niet prettig om te ontvangen? Waarom?
- Vertrouwt u de suggesties die het systeem u geeft? Heeft u het gevoel dat de aanwijzingen die het systeem u biedt, gebaseerd zijn op een wetenschappelijke basis?
- Wat vindt u van de modaliteit (ECA vs. Tekst)?

Privacy

- Welke omgeving factoren denkt u dat er meegenomen worden in de beredeneringen van het systeem?
- Wat vindt u ervan dat het systeem uw activiteiten registreert?
- Wat vindt u ervan dat het systeem uw medicatie inname registreert?
- Welke overige activiteiten denkt u dat er nog meer geregistreerd worden? Hoe denkt u hierover?

Overig

- Diversiteit teksten?