A virtual reality test of the usability of an artificial pancreas with diabetes 1 patients as co-creators

Author: Ronja Hüsgen
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT
The number of diabetes type 1 patients is growing and there is an increased need for innovation in treatment methods. Inreda BV has therefore started to develop the artificial pancreas which is a promising solution to enable patients to keep their lifestyles and prevent complication without being constrained by their disease. Prototypes of the artificial pancreas have been tested already, but have been found to be not very user friendly. Inreda BV therefore designed a new model with increased user friendliness and smaller and lighter hardware. The user interface of the new model Inreda BV designed still needs to be tested in terms of usability. It is also of Inreda BV's interest to find out how users cope with alarms that arise if something is not right with the device or the patient. This research is therefore aiming at finding a method for testing the usability of an artificial pancreas in a virtual reality environment involving the end-users as co-creators. In this paper different methods of usability testing, co-creation and virtual reality usage will be evaluated according to how suitable they are in this specific case and the best methods will be embedded in a test design approach in the VRLab of the University of Twente.

Supervisors:
Tamara Oukes (1st supervisor)
Ariane van Reasfeldmeijer (2nd supervisor)

Keywords
Usability, Co-creation, Virtual Reality, Artificial Pancreas, User involvement.

6th IBA Bachelor Thesis Conference, November 5th, 2015, Enschede, The Netherlands.

1. INTRODUCTION

Currently, around one million people in the Netherlands suffer from diabetes and numbers are growing. Diabetes does not only mean a change of lifestyle, but also the risk of complications like strokes, kidney disease, heart disease, visual impairment etc. for every single patient, and next to that it also means high costs for the state due to a loss of productivity of the workforce and treatment costs. The costs that occur through diabetes patients go up to billions of euros on a yearly basis. (Booz & Co, 2011)

Therefore there is a need for innovation of diabetes type 1 treatments that prevent complications and keep costs low, while being easy to use and creating a better lifestyle for patients. To invent such a solution is very costly, though. It would need to be tested by professionals, as well as by users to find out if it really makes it possible to keep a normal lifestyle and if the solution is reliable.

At this point, virtual reality (VR) prototype testing becomes an interesting opportunity for developers. Instead of creating and testing several generations of prototypes, what would mean very high cost and development time; VR could be a promising solution.

Booz&co (2011) suggest that the funding for innovations could be partly by insurances, which profit from new ways of treatment, because they would save a lot of money if medical complications are delayed or prevented, and another part should be paid by the government, which would save money in case of efficient solutions, because less diabetes patients would have to retire earlier due to complications and receive pension from the government.

Even though the reasoning sounds legitimate, it is difficult to raise money for research and development of new solutions or convince insurances to finance the projects partly.

The closed loop system by (among others) Inreda BV which consists of a subcutaneous insulin infusion (CSII), a continuous glucose monitor and a glucose control algorithm (van Bon et al. 2010) has already gone through prototype testing and is found to be a promising solution, but still needs some development in terms of reliability and very importantly usability. The subcutaneous insulin infusion was added a subcutaneous glucagon pump as a bi-hormonal approach to avoid a lack of glucagon in certain situations, where too much insulin is injected or the patient has a higher need of glucagon, such as in physical activity. This development might be possible to be realized with a virtual reality approach to keep the costs low, but still get an insight of the patients' latent needs in terms of usability, due to the testing which is close to the real experience of the product.

The end-users of the artificial pancreas are, as said above, diabetes 1 patients. Of course not every diabetes patient shares the same lifestyle. Some patients might be very active and enjoy working out several times a week and others might be rather calm, working a desk-job and only be physically active if necessary. Similar differences can be seen in eating habits of patients. Some individuals are very strict in following an appropriate diet to their dysfunction and try to keep a healthy nutrition, whereas other patients are not very convinced that reducing the comfort of their eating habits can have a big impact on their health. Furthermore the technical affinity of patients will vary. Those different kinds of lifestyles create a big challenge to develop a uniform product for diabetes treatment that can cope with all those different situations. The approach of Inreda BV with a control algorithm that is able to learn from actions the patient takes, combined with an easy to use interface, might be the answer to the problem of versatile end-users.

The virtual reality testing of the artificial pancreas would therefore need to include users from all different kinds of backgrounds to ensure that the system is satisfying in any case.

The feedback of the end-users after testing the product in a VR environment will be highly relevant for the development process of the product, because it enables the users to discover their own needs and desires through trial and error and afterwards might be able to communicate their latent needs (Füller & Matzler, 2007). Latent needs might also be observed by experts during the experiment according to Manon Spin and Roy Damgrave (Appendix D and E).

The existing body of literature gives detailed explanations on what usability is and which elements define good or bad usability, as well as various approaches on using virtual reality as a tool for testing or displaying in different sectors. There is also a decent amount of literature on co-creation involving the end-users of products, but there is no literature on combining those three theories for value creation. This research aims to address this gap in literature.

To test the usability of a product, valuable feedback from users is needed by the developing company to minimize risk and bring the best possible value to the market. This feedback can be achieved by using customer co-creation. Co-creation then is possible by letting the user experience the product. If there is no physical, produced product to test and limited funding to produce several generations of prototypes until it is ready to go to the market, the best possible solution is VR testing. This will keep the cost low and uncover customer latent needs during the test phase. The usability of a product can best be tested in (close to) real conditions, meaning that the customer can experience, see, hear, and try the product with as many senses as possible (Kuhlen and Dohle, 1995), which can be realized in a virtual reality (laboratory) testing. The three concepts put together should lead to the development of a satisfying product for both, developing company and end-user.

Therefore the following question should be researched: How can the usability of an artificial pancreas valuably be tested in a virtual reality environment with the involvement of diabetes 1 patients of different backgrounds as co-creators?

To do that, in this paper it will firstly be outlined what is already known about the testing of the usability of products, the use of virtual reality and the inclusion of customers in the development process as co-creators. Afterwards the key concepts and their relationship will be discussed and other fields of use will be outlined. The paper will be finished off by discussing and developing methods for testing and a design approach for a test for later studies as a conclusion to the research question.

1. THEORY

1.1.1 Introduction

The current literature delivers an adequate body of information on the usability of products. The usability of a product is nowadays one of the most important factors concerned in a purchasing decision (Dumas and Redish, 1994; Han et al., 2000; Babbar et al., 2002). This is due to the fact that the usability includes an emotional factor (Logan, 1994) and determines the ease of use (Han et al., 2000). A customer that purchases a product expects it to function well and meet their desires and needs (Babbar et al., 2002).

Usability is therefore especially important in the medical sector, where the characteristics of usability suddenly become vital factors for patients. Certainly a medical product that has innovative functions does not bring any value if its usability brings complications for the end-users, because they simply have problems interacting with the product.

Due to this reason it is only logical that direct feedback of end-users can bring high value to the developing company, in terms of clear knowledge of needs and wants of the customer, reliability of the product in test and therefore risk reduction, and a positive association of users with the product and brand.

On the one hand it can positively influence the purchasing decisions of users and on the other hand it means a reduction of risk for the developing company. The risk of developing and producing a product, bring it to the market and experience unsuccessful sales, due to unmet needs on the customer side, can be narrowed (Prahalad and Ramaswamy, 2004).

On the customer side co-creation means being heard and communicating needs that are sometimes even unknown before being involved in the development process (Füller and Matzler, 2007).

Due to the reason that users have latent needs (hitherto unknown), it would be necessary to let the users try the new product to figure out their needs, wants or preferences. The problem with that is that developing prototypes to have them tested and possibly not satisfactory (so prototype a new generation) is very costly and therefore brings risk.

Virtual reality (VR) makes it possible, though, to enable users to try the product before producing it. VR usually consists of a "computer-generated 3D environment – called a 'virtual environment' (VE) – that one can navigate and possibly interact with, resulting in real-time simulation of one or more of the user's senses" (Guttentag, 2010). The interaction with the VE is operated with a specific input device that allows the manipulation of objects (Kuhlen and Dohle, 1995).

1.1.2 Usability

The three key concepts discussed in this paper are usability, co-creation and virtual reality. The first one, usability, is defined by the following characteristics. As named above, usability is nowadays considered as one of the most important factors in a person's purchasing decision (Dumas and Redish, 1994; Han et al., 2000; Babbar et al., 2002). It is a generic term for ergonomic product quality that is used to replace other terms like user-friendliness or ease of use (Dzida, 1995; Babbar et al., 2002). The main purpose of usability is to describe whether a product meets the needs of a user and fits with its work practices or activities (Bevan, 1999; Babbar et al., 2002). "The ISO 9126 Standard defines usability as a set of attributes that on the effort needed for use, and on the individual assessment of such use, by a stated or implied set of users" (Bevan, 1999; Babbar et al., 2002). The scope of usability was later expanded by the ISO CD 9241-11 (1998) to include effectiveness, efficiency and satisfaction with which users could achieve a specified goal with the product (Babbar et al., 2002). This set of attributes that is defined by the ISO standards is going to be the main definition of usability in this paper, due to practical reasons of using the outcomes of the study for evaluations at Inreda BV. The definition is expanded and supported by the other attributes named. "Effectiveness is defined here as the accuracy and completeness with which users achieve specified goals in particular environments. Efficiency refers to the accuracy and completeness of goals achieved in relation to resources expended, while satisfaction is defined as the comfort and acceptability of using a system." (Babbar et al., 2002)

Schneiderman (1992), Hix and Hartson (1993) and Nielson (1993) also add the factor of ease of learning, memorability, error rates and preferences to the list of defining characteristics.

Han et al. (2000) propose a different approach on defining usability, though. The authors state that the dimensions of usability can be separated in two groups. The first groups, named the performance dimensions, measure the performance of the user. Those are broken down into three categories: perception/cognition, learning/memorization, and control/action. Those categories hold 23 performance dimensions in total.

The second group is called the image/impression dimension and is again divided in three categories: basic sense, description of image, and evaluative feeling/attitude. This group holds 25 performance dimensions in total.

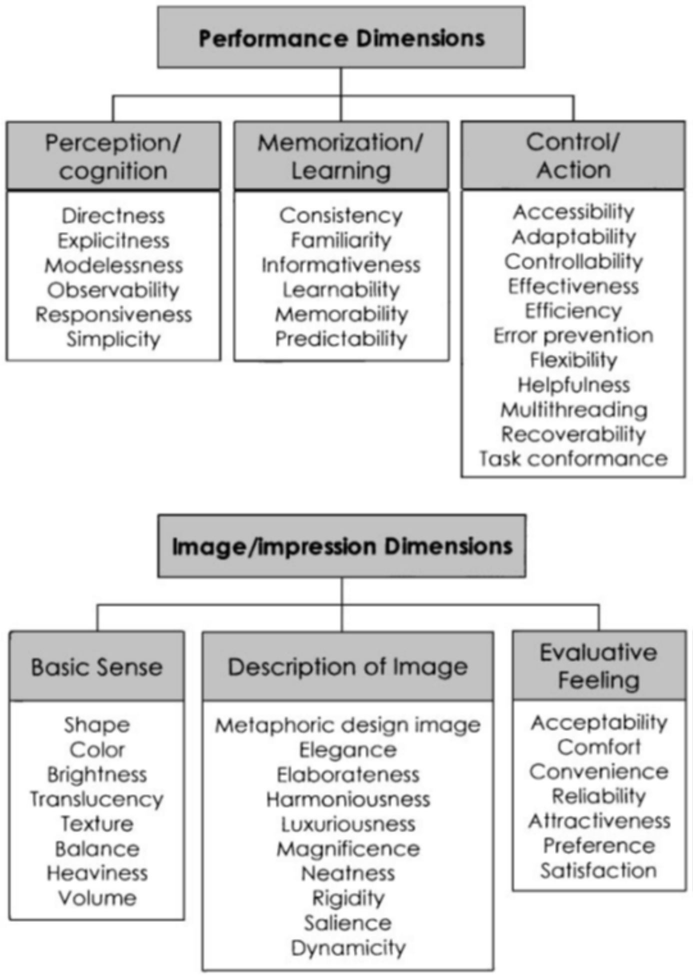


Figure 1. Usability dimensions (retrieved from Han et al., 2000)

With this scheme it is possible to clearly define if a product has a high or low usability for a specific user and get a better understanding of what is important to users, but to get to that knowledge it is important to involve the user in the testing somehow and this is where co-creation becomes a possible solution.

1.1.3 Co-creation

The alignment with potential customers is crucial for business success and associates an understanding of which product attributes are important to them, next to the incorporation of feedback from customers (Whiteley, 1991; Veryzer, 1998; Cristiano et al., 2000).

The involvement of the customer in the development process is also known as co-creation. Nowadays the value creation process is shifting from a product- and firm-centric view towards a customer-centric view (Prahalad and Ramaswamy, 2004). In the earlier days value exchange and extraction were the main functions performed by the market, which was separated completely from the value creation process; therefore the only point of communication was between the firm and the customer on the market (Prahalad and Ramaswamy, 2004). In the last decades this image shifted towards a more interactive image. Dialog, access, risk-benefits and transparency (DART) are the building blocks of the new function of the market (Prahalad and Ramaswamy, 2004). This new value creating market includes customer-to-customer communication, consumer-company interaction and therefore welcomes a co-creation approach.

Co-creation is already used in different intensities and different industries. The term includes a wide range of activities. It can describe the possibility for customers to give feedback and ideas to a company on online platforms, and go as far as involving selected users in the development process of prototypes or the co-creation of a person with a disease in finding the right treatment together with medical staff (Prahalad and Ramaswamy, 2004).

In the case of this paper, where products of the medical sector should be tested and developed, communication and transparency are especially valuable to create trust and satisfaction among patients. “The health provider-patient relationship has traditionally been asymmetric, with the power in favor of the provider” (McColl-Kennedy et al., 2009). This situation is changing through co-creation in the medical sector and is more balanced, since the patient is gaining the power to give feedback, search for different providers, get insights and is needed by the provider to gain the necessary feedback to become better.

The problem with customer co-creation is that customers often cannot communicate their needs and wants, because they are latent needs and first need to be discovered (Füller and Matzler, 2007). Figure 2 illustrates the users’ problems to communicate their needs with the Kano model.

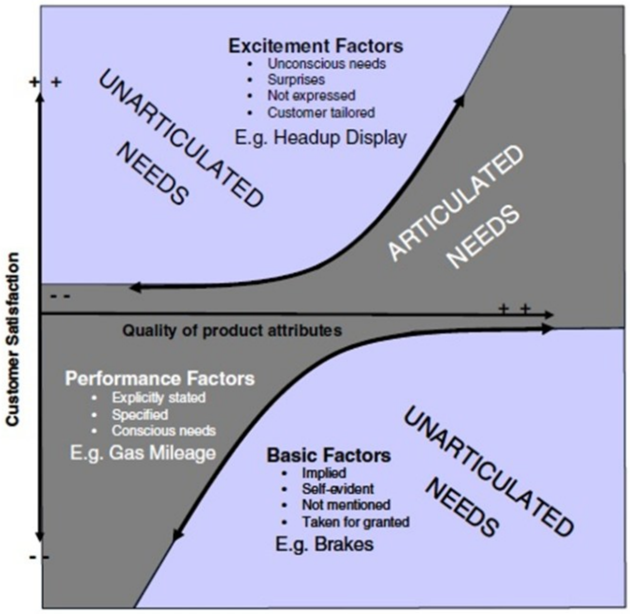


Figure 2. Customers' problems to articulate their needs illustrated by the Kano model (retrieved from Füller and Matzler, 2007)

Latent needs can in many cases be discovered while testing a specific product in its specific environment. The testing of the product often gives the user ideas about which characteristics are missing, dissatisfying or working well and therefore also helps to identify characteristics which nobody thought of before trying the product in use.

As it was explained before, the development of several prototype generations for trial and error tests is very costly and funding for medical research is limited, therefore virtual reality (VR) testing can be a good and innovative alternative.

1.1.4 Virtual reality

The first traces of virtual reality go back to the 1960s, but it was not until the 1980s that VR started to be seen as an opportunity for research centers and industry. In 1990 VR systems were firstly introduced to the research community and have now found many application areas (Cruz-Neira, 1998).

VR aims at addressing as many senses as possible (Kuhlen and Dohle, 1995) and can be classified by its ability to provide physical immersion and psychological presence (Gutiérrez et al., 2008; Guttentag, 2010). Immersion in this context refers to the degree of isolation of the user from the real world (Guttentag, 2010). “In a ‘fully immersive system’ the user is completely encompassed by the VE and has no interaction with the real world, while in a ‘semi-immersive’ or ‘non-immersive system’ […] the user retains some contact with the real world” (Gutiérrez et al., 2008; Guttentag, 2010).

To make the illusion of the virtual reality test successful and enable the user to dive into the virtual world with high immersion, it is very important that the responsive virtual environment reacts in real-time to actions taken by the user (Blach et al., 1998).

The VR input devices that help to interact with the 3D animated pictures have various possibilities and modern devices can be operated by speech, gesture, sound, position, touch and many more (Blach et al., 1998).

There are different kinds of virtual reality applications nowadays. Virtual reality differs from “second life”, where people interact in a virtual world on the internet through avatars, to high tech laboratories for research and testing. Many companies across industries operate open labs on the internet where customers can co-create products and give feedback (Leminen et al., 2012). The computer gaming industry also works with virtual realities where gamers around the world can join a virtual world and play with or against each other through avatars.

Virtual Reality also finds application many different fields. In marketing, for example, it is used to display products on the internet. Customers can then see the product in a 360° display, click on functions to try them or listen to it. Studies have shown that the 3D animated products enhance customer learning compared to 2D models (Li et al., 2001, 2002, 2003; Suh and Lee, 2005). This application would be an example for non-immersive VR.

Another example of VR application in different fields would be the education sector. VR gives the opportunity to make learning more vivid and also helps to improve the ability of students to analyze and solve problems (Pan et al., 2006). Students can for example virtually walk through historical events with an avatar instead of just reading about it. Cruz-Neira (1998) also explains the potential of VR in museums. Historical events could be recreated with evidence from the past and a VR environment could be created for the visitor.

The application of VR could also be interesting for the tourism sector, as explained by Guttentag (2010).

Another interesting VR practice is to use innovative systems to help disabled people to complete tasks where they usually would be hindered by their disabilities or train body functions that are disabled to function better again (Kuhlen and Dohle, 1995; Di Gironimo et al., 2013).

Virtual reality has various application fields and gives room for innovation in the future.

To test whether virtual reality is a useful tool in this case a summary of all attributes of a SWOT analysis (Strengths, Weaknesses, Opportunities and Threats), which have been named in this paper so far, of virtual reality testing in this specific case will be conducted. A SWOT analysis is originally used to summarize the Strengths, Weaknesses, Opportunities and Threats of a company to determine its position on the market, but is nowadays used in any kind of decision making in various fields next to the business sector (Rizzo and Kim, 2005). The following table (Table 1) will summarize the main strengths, weaknesses, opportunities and threats in the case of using virtual reality for usability testing involving end-users as co-creators.

Table 1 SWOT analysis of VR testing in the specific case of this paper

Strengths	**Weaknesses**	**Opportunities**	**Threats**
VR enables tests that are close to tests in the reality	The tests are “only” close to tests in the reality	VR technologies are a constantly innovating sector and promise a growing variety of application	The failure of the created virtual world to appear real (According to Carrozzo and Lacquaniti (1998) certain conditions need to be fulfilled to make a computer-animated world appear realistic to the subject group)
It reduces the risk for the developing company, that would occur if the product would be produced and would fail on the market (Prahalad and Ramaswamy, 2004)	Patients might react differently in a VR environment than in a real life experience, because the laboratory environment is new to them (Appendix E)	VR is usable in many different kinds of tests (not only the usability can be tested, but also for example the value-in-use or the reliability of functions)	
It safes time compared to the process of producing prototypes and testing them in a real life scenario (whole days, compared to day segments in VR), as in van Bon et al. (2010, 2012, 2014) where clinical tests were conducted and took several days		The conducted tests can be repeated in same/similar conditions with same or different co-creators (in real life it is hard to ensure similar conditions)	
VR involves less risk for patients that are involved in the tests of a medical product (If the product is insufficient it could harm the co-creator)			

According to the SWOT analysis in Table 2, the strengths and opportunities of using virtual reality in this case are distinctively stronger. The named threat can be diminished by creating the VR environment in a professional way, with modern technology and the named weaknesses can be reduced by trying to make the co-creators comfortable with the laboratory environment and let them adapt to the surrounding properly.

1.1.5 Usability and Co-creation

As it was described above usability can best be tested by somehow involving the end-user. Feedback of users is needed to come to a conclusion of whether or not a product has a satisfying usability. Co-creation can deliver this valuable feedback for the developing company and can help to reduce risk. If the company would bring a product to the market that turns out to have an unsatisfying usability, users will be frustrated (Babbar et al., 2002) and refuse to purchase the product and the developing company could experience great losses (Prahalad and Ramaswamy, 2004). Sometimes needs or wants of customers are simply unknown to the developing company and co-creators can help to communicate those needs and wants.

The combination of usability and co-creation can already be found in the existing literature, but there is still a problem when it comes to the testing of products involving co-creators, because for those tests prototypes would be needed and producing a prototype, having it tested with a trial and error approach by co-creators, and then creating a new generation of prototypes to have it tested, is a very costly process.

1.1.6 Co-creation and virtual reality

The combination of VR and co-creation can already be found in different applications, for example in online marketing displays where customers can see a virtual version of a product and interact with it and write comments on it or are asked to answer to questionnaires (Suh and Lee, 2005).

VR helps to keep the costs of creating prototypes low, because designs of products only need to be changed virtually after generating customer feedback.

The existing literature on co-creation and virtual reality mainly focusses on VR in online environments (non-immersive VR) like company websites or second-life and involves users only to a small degree, as in comments or questionnaires after interacting with the virtual model of the product. The large number of users that give feedback to the company in online testing also lets the user's opinion be one in a large group; whereas the user could be more involved if a sample of users would be collected that represents the whole "population" of users. This type of non-immersive co-creation in VR can also be seen as a Think-Tank that helps the company to get an overview of customer needs (Leminen et al., 2012).

VR also gives the developing company together with the user the opportunity to discover latent needs of users (Füller and Matzler, 2007), as discussed earlier.

Due to the light involvement of co-creators in existing VR tests, those methods are not applicable in usability testing, since the users would need to try the product in more depth to build an actual opinion on it.

1.1.7 Virtual reality and usability

To try a product in depth to make an opinion about its usability becomes possible in a virtual reality laboratory. A virtual reality laboratory contains the latest technology to let an item or environments appear very realistic, addressing different senses of the human body at the same time (Kuhlen and Dohle, 1995).

This realistic image of a product in a virtual reality laboratory can then be tested in a specific scene using different input devices to allow interaction. Di Gironimo et al. (2013) use a virtual reality environment to test the usability of a wheelchair-mounted robot manipulator. The testing is conducted at a rather late stage of the development process, though, different from the test that should be conducted according to this research paper.

The existing literature on virtual reality and usability shows that usability can very well be tested in a virtual reality environment. It is necessary, though to involve users in the testing to get to valuable feedback, because only users for which the product is determined to be (in the case of this study, diabetes type 1 patients) can evaluate if the usability is satisfactory for their specific situation.

1.1.8 Usability, co-creation and virtual reality

Considering the discussion of existing literature and the study which should be conducted according to the outcomes of this paper, a combination of usability testing in a virtual reality laboratory involving users as co-creators is the solution to come to a satisfying end-product.

The relationship between the concepts named above is shown in Figure 3.

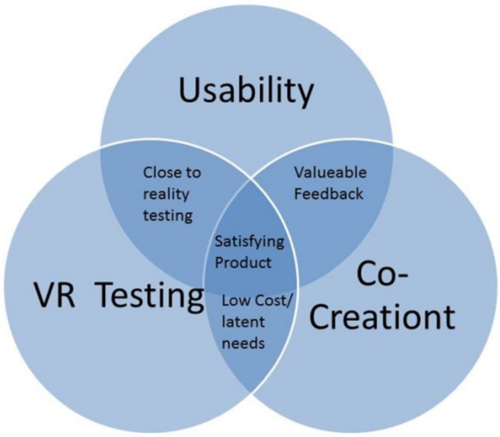


Figure 3. The relationships of the key concepts

1.1.9 Conclusion

There is a solid body of literature on each of the key concept separately, but there is a gap when it comes to combining those concepts to an approach to test usability in a virtual reality environment involving end-users as co-creators in an early stage of the development process. It is possible to find the combination of VR and co-creation, like in online marketing displays, but to test a product on its usability in a fully or semi immersive system and involving customers in the process to gain valuable feedback is still a gap in literature. The combination of those three theories can, especially in the sector of medical products be very promising, because in this field it is highly important to develop close to the users' needs, even though they might be latent needs, to ensure good treatments.

To combine those three concepts in this case it would need to be further tested, how a testing experiment for the artificial pancreas could be designed. Which VR methods would be most applicable, how many patients should be involved in the testing to receive valuable feedback, with which methods will the test outcomes be analyzed.

To fill this existing gap in the literature, different methods for analysis and testing will be introduced and concluded with a design approach for the VR testing. This study is meant to lay the foundations for further studies and the conduction of the laboratory tests.

Furthermore the combination of usability, co-creation and VR testing will open the door for further research studies to apply the concept in different fields.

2. METHODS

2.1 Methods for involving users as co-creators

When evaluating methods for testing the usability of a product involving the end-users in the development process as co-creators, it is necessary to firstly define the different methods of involving users as co-creators. Users can be involved in different stages of the development process as well as with different degrees of involvement.

The involvement of users in an early stage of the development process usually brings a higher involvement of the user, because the user has more influence on the design of the product. If users are involved in a later stage of the development process, the influence of the users is usually smaller, because the product is already designed and prototypes have been designed with certain attributes. The developers are therefore less open for changes and the tests are the last hurdle and insurance before going to the market. The user involvement in early stages of the development process is also more efficient, because costs would be involved to make changes in the product design at later stages in the process (Ehrlich and Rohn, 1994; Noyes et al., 1996; cited by Kujala, 2003).

The involvement of co-creators and the development stage at which users are involved are therefore intertwined.

The term co-creation is somehow brought and can be used for different forms of user involvement. The degree to which a user is involved in the development process varies between light and high involvement as shown in the following examples.

2.1.1 Light involvement of co-creators (late stage of development process)

Prahalad and Ramaswamy's research (2004) on co-creation experiences, where they name the example of cancer patients who are nowadays able to inform themselves via internet or other sources and design their specific treatment methods in cooperation with the medical staff, is an example of light involvement of users. This kind of user involvement does not give the patient the opportunity to fully create something new, but it does make it possible to intervene if a certain treatment does not concur with what the patient read, heard or knows and make suggestions for different treatments that fit his or her circumstances best.

Another example of this light involvement of users in co-creation is the user involvement in the design process of already existing products like shoes at Nike or cars at Mercedes, where customers can select from several options to put their favorite parts together in one customized end product (Füller and Matzler, 2007). This case is an example of mass-customization. While there is only a few selected people involved in some co-creation processes, in the last two examples it is possible to involve a wide range of customers, keeping in mind that the first example stands for (cancer) patients in general who have the possibility to co-create their treatments.

2.1.2 Medium involvement of co-creators (all stages of development process)

A slightly more involved co-creation method is introduced in Kohler et al.'s research on avatar-based innovation (2011). In this study the researchers discuss the use of second life as a source for co-creation. The authors state that according to a study by von Hippel (2001), traditional market research does not clearly identify the customers' needs and wants anymore. In a second life experience users come together in a virtual world, a computer-generated 3D environment, and interact through avatars. An avatar is the virtual representation of an individual user, which can be manipulated in the virtual world and makes it possible to interact with other users' avatars (Castranova, 2005; cited by Kohler et al., 2011). Companies started to use those virtual worlds to let selected users experience the prototypes of their products through their avatars and gain direct feedback (Kohler et al., 2011). In this type of co-creation the users take an active role and have the chance to co-create value together with the developing company (Prahalad and Ramaswamy, 2004; cited by Kohler et al., 2011). The number of involved users here can also be relatively high, but might also be small according to the selection made by the developing company. A difference to the first example of co-creation is, that in this case the company can select the users it is co-creating with compared to the doctor-patient relationship that was described earlier, where the doctor has to co-create with any patient that is in need of treatments. The co-creators can be involved in early and late stages of the development process, because the products in test are virtual and can be a concept that has only been designed, not developed yet, and virtually created, but it can also be a virtual construction of an existing product and therefore involve users at a late stage in the development process.

2.1.3 High involvement of co-creators (late stage of the development process in this example)

The third co-creation approach is very user-focused and involves customers to a very high degree. Di Gironimo et al. (2013) conducted a case study to test the usability of a wheelchair-mounted robot manipulator and involved patients in the prototype testing. This type of user involvement can also be seen as co-creation and involves specifically selected users that represent all patients that could be using the new technique. The number of co-creators was therefore relatively low compared to the methods introduced before. The testing was done in a virtual reality environment, using the prototyped product to manipulate objects virtually. Therefore in this study the patients were involved in a later stage of the development process, where two different kinds of prototypes had been developed and the test was supposed to give an indication about which control device would be more suitable. Due to the fact that the number of co-creators is limited in this case and evaluations and analysis of the generated feedback are rather extensive, the degree of involvement of the user is very high.

2.1.4 Conclusion of co-creation methods

In the case of this research paper where a method is searched to test the usability of an artificial pancreas in a virtual reality laboratory, involving end-users as co-creators, the last method named is most applicable, because as in this method our study is about a medical product, which asks for a very user-focused approach as it was discussed before. This method involves a carefully selected sample of patients as co-creators, which is also very applicable in this case, because the end-product should be easy to use for people with all kinds of characteristics. As in the example named above, a carefully selected set of users will be asked to join the co-creation process and will be involved to a very high degree. In this case the method will be different in terms of the stage in the development process where the users will be involved. In Di Gironimo et al. (2013) the co-creators were asked to test a finished product, whereas in this study, the users will be asked to try out a virtual design of a new model of the product, which lets them be involved very early in the development process. The testing of the device will be fully virtual (without a physical prototype), though, because the users will still be involved in the design process of the user interface. Co-creators involved will get the chance to freely express their opinions and feelings about the product and therefore carry more responsibility, as well has have more opportunities than in comparable experiments.

2.2 Methods selection for assessing usability

To test the usability of a product it is not only important to design a test environment, but also to design a method to collect feedback from the involved users as well as from the developers, to analyze and evaluate it in order to create value for the developing company.

As earlier studies show, qualitative methods are most appropriate for consumer research (Van Kleef et al., 2005). Qualitative analysis helps to make sense of social observations (Babbie, 2010) and helps the person conducting the test to clearly formulate or express his or her answers. Quantitative answers on the other hand, help to make observations more explicit and make it easier to aggregate, compare and summarize data (Babbie, 2010). Therefore a mixed approach of qualitative and quantitative methods would be most applicable for this study, because this study aims at designing a test that involves social observations (in the sample group of patients) while explicit data is needed to compare test results of different participants and make the test repeatable at a later point of time. To select the right method(s) for the evaluation of user involvement, different methods will be introduced and discussed in the following.

As it was described earlier according to the ISO (1998) usability consists of three main attributes, which will lay the basis for this usability test and therefore also the evaluation of methods:

2.2.1 Effectiveness

The effectiveness of a product in the experiment can be tested by the expert that monitors the interaction of the user and the product by counting the amount of tries to complete a task and the amount of completed tasks. With those numbers it is possible to calculate a success ratio and compare those ratios or relate them to certain patient characteristics.

2.2.2 Efficiency

The efficiency of a patient using the product could be measured by taking the time the patient needs to complete each task. This quantitative data can also be compared or related to other characteristics as in the case of the effectiveness outcomes.

It might be interesting to see how times vary between users with high and low technical skills or older and younger patients.

2.2.3 Satisfaction

Measuring the satisfaction is slightly more complex. The satisfaction of a patient with the product is a subjective feeling and can only be communicated by the patient. Therefore the proposition in this paper is to create a direct questionnaire according to the situation that the co-creator undergoes during the test which should be answered by the patient after the virtual reality testing. The experts are then able to evaluate the questionnaires by transforming parts of the qualitative data to quantitative data and retrieve measurable data from the direct feedback of the patients and compare the pre- and post-test outcomes with one another.

As an addition some questions should be qualitative and the patients should have the opportunity to discuss any additional ideas, comments or feelings at the end of the questionnaire to ensure a high freedom of expression and the opportunity for latent needs to be discovered.

To test the usability of an artificial pancreas according to this definition of usability, this paper proposes four different techniques from a number of existing methods available in the literature and evaluates them on the basis of five criteria which are partly retrieved from existing literature and partly created to fit this specific testing, to bring the planned experiment to the desired outcome.

To introduce the evaluation of methods, the following table (Table 2) will show the methods and its advantages and disadvantages regarding the evaluation criteria.

Table 2 Method evaluation

		Focus Group Interviewing	Direct Survey Questionnaire	Importance Ratings	Observation through Experts
Degree of Involvement	High	X	X	X	X
	Low		X		
Product Familiarity	High	X	X	X	X
	Low		X		X
Freedom of Expression	High		X		X
	Low	X		X	
Validity and Repeatability	High		X		X
	Low	X		X	
Applicability in VR environment	High	X	X	X	X
	Low				

The degree of involvement describes the degree to which the user is involved in the co-creation process. As described in 3.1. with the examples of different cases, co-creators can be involved in the development process to different degrees and at different stages according to the product and type of test.

The product familiarity describes the degree to which the user is familiar with the product that should be tested before the experiment starts (Eliashberg, 1997; cited by Klaver, 2015).

Furthermore in this paper the "freedom of expression", the "validity and repeatability", as well the "application in a VR environment" were added to evaluate the methods in the specific case of this experiment.

The freedom of expression is concerned with how freely the user can express his/her thoughts, feelings, emotions, and ideas etc. about the product in the evaluated method. As discussed above the study this paper prepares does not only try to evaluate the existing attributes of the product, but also tries to detect latent needs. Therefore it is very important that the co-creator can freely express his/her feelings and thoughts, because he/she is not aware of these needs him-/herself, yet and it might be possible to retrieve latent needs from those freely spoken thoughts.

The validity describes in how far the outcome of the test is relevant for all diabetes patients (in this case) rather than only for certain individuals involved in the test. As Babbie (2010) describes it: "validity refers to the extent to which an empirical measure adequately reflects the real meaning of the concept under consideration". The repeatability aspect describes in how far the test would bring the same outcome if it was repeated at a different point in time with different participants. This criterion is important, because the test might have to be done several times if changes are made in the design of the artificial pancreas and a new model is introduced to new co-creators, the test needs to be repeatable to compare which of the models had a better usability.

The Application in a VR environment is of high importance in this paper, because it is supposed to lay the foundation of the later conducted experiment in a virtual reality laboratory. Therefore all methods that are selected need to be applicable in such an environment.

2.2.4 Focus Group interviewing

In this method, potential customers are introduced to the product or concept at hand in small groups and are asked to discuss the value the product would bring to them. This qualitative approach helps to better understand what is important for the customer and get a first impression of the reactions on the product or concept (Anderson and Narus, 2004). The discussion between the users is open-ended and unstructured (Calder, 1977; cited by Klaver 2015), whilst a researcher acts as a moderator to keep the discussion on topic and make sure that it comes to an outcome in the end (Anderson and Narus, 2004).

Focus group interviewing gives the end-user, in this case the patient, a high degree of involvement because it enables the participants in the group to exchange experiences with the product in test (if the product was tested by them before the discussion), first impressions or feelings with the other patients. To enable such a discussion, the participants would need some product familiarity to have exchange their experiences. The problem with this method and the application in the medical sector is that patients would need to share sensitive information in the group, which could prevent them of expressing their true feelings and experiences of the product or their medical history with other participants. This is a large limitation in order to gain feedback that can be used to improve the usability of the product. Furthermore Calder (1977; cited by Klaver, 2015) argues that focus group interviewing is not very objective. A similar discussion with other participants and other moderator could bring a different course of discussion and different outcome. The method is therefore not valid and repeatable. Focus group interviewing can be applicable in a virtual reality environment, if the discussion should be held after participants try the product. If the product is unknown and the discussion is held for a first impression discussion it is not very applicable in a VR environment. Due to the reasons named this method is not very suitable for the experiment prepared by this paper.

2.2.5 Direct survey questions

Direct survey questions are used especially often to find out how a specific group of people thinks about a certain topic, product or event. In the article of Business Insights Ltd. (2011) about "The Diabetes Device Outlook to 2016", for example, a group of Doctors answered a direct survey questionnaire about a specific product for diabetes care to give an estimation about the value this product could bring to the market.

Direct survey questions are not only used for the value assessment of existing products, but are also applicable in the development of new products. In the medical sector it is very important to involve the end-user in the development process and get insights about feelings, emotions, usability factors or abilities of end users, as named before. Due to that it is very useful to determine the value of medical devices for specific groups of people.

In their research study about a virtual reality approach in the testing of a wheelchair-mounted robot manipulator Di Gironimo et al. (2013) used a combination of different value assessment methods at different stages of their experiment. They combined a preliminary direct questionnaire, which was mainly to select the right focus group, with a focus group value assessment to test the new product concept in the virtual reality laboratory. This method can therefore not only be used to evaluate a product after an experiment, but also to select a group of co-creators.

Direct survey questions have the advantage, compared to other methods, that qualitative data can be transformed to quantitative data and therefore make it possible to put a specific value on an asset. Due to the reason that the qualitative output of the questionnaire becomes measurable in quantitative values, this method is also valid and repeatable. It is also possible to combine questions that are transformed to quantitative data and questions that give qualitative output, to ensure that the co-creator can express what is on his or her mind freely.

The patient can have a high or low degree of involvement. In the case of selection of a focus group, the involvement of the respondent is relatively low as he or she was not exposed to the product yet and does not have much information. In the case of evaluation through a direct questionnaire, the user has a high degree of involvement, can express his or her feelings freely in the answers and is familiar with the product. Therefore the user can also have a high or low product familiarity depending on the kind of direct questionnaire. Often a combination of different kinds of questionnaires is applied before the user is familiar with a product and then after the user has gained product familiarity.

The method of direct survey questions therefore seems very suitable for the planned usability testing, both for selecting the right focus group and to enable the patients to put their feedback into words. It is also very applicable in a virtual reality environment, as tested by Di Gironimo et al. (2013) firstly to select the right focus group for the VR experiment and secondly to receive an assessment of the product by the co-creator.

2.2.6 Importance rating

In an importance rating patients are introduced to a product and can rate product attributes according to their personal feeling of importance of that specific attribute. In the study of van Bon (2014) for example, participants said that they were very satisfied with the product accept its size and format. With an importance rating the developing company could now find out whether the characteristic of size would be more important to the participants than other attributes.

Even though this method is very applicable in the medical sector in general and would also be interesting in the case of the artificial pancreas usability study, importance ratings do not give the participants enough freedom for expression that is needed for this study. The participants are provided with a certain set of attributes to rate, but this study aims to also find out about hitherto unknown product attributes that could solve the patients' latent needs. Furthermore this method is not valid and repeatable, because it is very subjective to rate attributes. It would need a very large number of participants to gain quantitative data that is statistically valid. It is also not necessarily applicable in the virtual reality environment, because the product attributes that are to be rated are known and can be named or shown without being put on a finished product and then rated. The virtual reality experiment would in this case only give an indication to the co-creator on how the attributes are put together in the end product, but is not crucial for the importance rating. The VR experiment could detect latent needs, but then fails to communicate those needs. Due to all the reasons named, this method is not suitable for this study.

2.2.7 Observation through experts

The last method that will be introduced is the observation through experts. When conducting a virtual reality experiment with patients as co-creators they will need guidance through the experiment by an expert. This expert will most likely be a person that is involved in the development process in the company, as in van Bon et al. (2010, 2012) where medical staff monitored and evaluated the performance of the artificial pancreas and the reactions of the patients.

The expert can observe the patient during the test and take notes on specific actions of the user during the test as an objective opinion on the patient's performance.

The specific actions and points of attention need to be clarified before the test and similarly conducted for every patient in test to make it repeatable in all of the tests in this experiment and ensure a reliable outcome.

The observation through experts gives the patient a high degree of involvement, because the user needs to fully interact with the product to make his or her performance observable and give the experts the opportunity to retrieve feedback out of the actions of the user. This feedback might sound passive, but is highly relevant and it is still possible for the patient to express satisfaction or dissatisfaction through actions, reactions or speaking out loud. The product familiarity in this case can be high or low, depending on what should be observed. It can be low if the user is supposed to figure out how to use the product him-/herself or it can be high if the user is supposed to complete specific tasks with the product which requite prior knowledge of the product.

The observation through experts can be used in any environment where a user interacts with a product. It is therefore highly applicable in a virtual reality environment.

In general this method is very helpful to gain objective feedback on the users performance with the product in a test environment and therefore on the usability. For this study it would be suitable in combinations with other methods that make the overall outcome valid and repeatable.

2.3 Methods chosen for assessing usability

It can be concluded from the previous evaluation of methods for assessing usability, that a direct survey questionnaire and the observation through experts are the most suitable methods for this research study, because they meet the most evaluation criteria and promise to give a detailed whilst explicit feedback from the co-creators if they are used in the right synergy.

To reach this synergy that ensures a statistically valid whilst informative and detailed outcome of the study, this paper proposes to use two different angles in the testing of usability according to the ISO definition. The first angle will be from the expert's point of view and the second angle will be from the patient's (Co-creator) point of view.

From the expert's point of view it will be possible to test the effectiveness and efficiency of the product with the use of observation through experts.

The Satisfaction of the patient with the product will be described from the patient's point of view using a direct questionnaire that consists of qualitative and quantitative questions, which will be analyzed by the experts.

Before the testing begins it would be appropriate to send out a direct questionnaire to prospective co-creators in order to find out if they are suitable for the experiment like in Di Gironimo et al. (2013). This is mainly important due to the reason that it is important to have a wide range of patients with different life-styles and skills in order to gain a reliable sample of outcomes.

2.4 Methods for using virtual reality

As well as for co-creation and usability there are different methods for the use of virtual reality environments. The term virtual reality describes many different forms of interaction between a person and a computer-animated item or world (Kuhlen and Dohle, 1995).

There are different kinds of visual techniques to create the virtual environment on a screen, as described in Carrozzo and Lacquantini (1998), as well as different manipulation (or input) devices to interact with the environment.

As described above, virtual reality starts with giving a customer the opportunity to view 3D pictures of a product online and click on features to try them out, see or listen to more information and move them around (Suh and Lee, 2005) and goes as far as creating a virtual world on big curved screens that let the user dive into the computer-created environment and interact with this virtual world through different input devices as in Cruz-Neira's research on "Making a virtual reality useful" (1998).

The input devices that can be used are constantly innovated and differ according to the task that needs to be fulfilled by the user. In online product displays, a common computer mouse can be used to interact with the virtual product. In laboratory tests on the other hand there are many different opportunities. Corrozzo and Lacquaniti (1998) for example introduce search-coil systems, which are tracking devices that return the position and orientation of a sensor in real-time, optoelectronic devices to turn 2D pictures into 3D, gloves that transmit hand movements to the computer-based environment, and force-feedback which is a similar approach as the today well known touch screens. A limitation of the explanation of input devices by Corrozzo and Lacquaniti is that their paper is already written in 1998 and the introduced techniques are not the most innovative on the market anymore, due to fast changing technology. More innovative technology will be introduced in the following section, when the equipment of the VRLab at the University of Twente will be explained.

The study design for testing the usability of an artificial pancreas, involving the end-users as co-creators, will be made according to the possibilities of the VRlab at the University of Twente. Therefore this paper will focus on the equipment that is available in this specific laboratory.

The virtual reality laboratory at the University of Twente (VRlab) offers various opportunities for testing with up-to-date equipment and tools. In this specific virtual reality laboratory it is possible to let participants dive into such a virtual world by bringing real motion to the screen. This possibility is very useful for prototype testing or building of (virtual) prototypes, user tests, benchmarking, collective mind mapping, and many more (VR-Lab, 2012).

The VRLab offers different techniques for displaying things on various screen techniques as well as different input devices to interact with the environment on the screen.

This paper will focus on the explanation of the most useful techniques for this specific research.

The most interesting screen techniques for this research are the “Oculus Rift”, the “Elumens” and the “Theatre projection screen”. All information about the equipment that is available in the VRLab is retrieved from the website of the laboratory (VR-Lab, 2012) and the information given by Roy Damgrave in an interview (Appendix E), who is among others developing the VRLab constantly.

The Oculus Rift is a virtual reality head-mounted display, which allows full immersion in the virtual world. It allows a 360° picture and the user can look around in the picture by turning his/her head like in the real world. The experience therefore becomes very realistic. It is even possible to see the movement of the users own hands in the virtual world. He/she can simply look at his/her hands and move them and will see the same movement as in the real world. The hands can therefore simply interact with the virtual environment and no additional input device is needed. It would even be possible to give the user a model version of the artificial pancreas and make it visible in the virtual world, so that the user can press the actual button in the real world, but would see the outcome of this action in the virtual world. Overall it would be a very promising technique to use in this case, but the problem would be that the Oculus Rift lets the user completely dive into the virtual world and does not allow a way out, by turning the users head. The device would need to be taken off to get back to the real world. This could be very confusing for people that have no experience with VR or the latest technique in general and could according to Roy Damgrave disturb the participants and make them feel uncomfortable, which could influence the outcomes of the study.

The Elumens in the VRLab also seem to be a suitable screen technique for this experiment. The device in question is about 150cm in diameter and operated by one computer and projector. It is a semi-immersive hemispherical display system that enables a 160° field of view, without the use of goggles, glasses or helmets. It is very suitable for simulations and can directly be connected to an input device. The problem with Elumens is that the user is sitting at a desk in front of the screen, which influences the freedom of the virtual experience. If the scene on the screen shows an avatar that is running and the user is sitting on a desk, the integration in the virtual environment is not as easy as if the user could move around freely. Another problem for the immersion with the virtual world is that the screen is rather small and the user can still see the laboratory around him/her. Therefore the Elumens is not the perfect solution, as well.

The Theatre projection screen is a large eight by three meters curved projection screen that has a projected image of two seamlessly blend dual projectors. Its purpose is to let one or more individuals emerge in a virtual environment. The unit in question is a semi-immersive system (Gutiérrez et al., 2008; Guttentag, 2010). The picture on the screen can be supported by a 7.1 sound system, which helps to make the virtual environment feel real. The screen is slightly curved, which allows the user to look around in a virtual environment and immerse in it. The Theatre Screen is very suitable for this study, because it can create a high degree of immersion with a virtual environment without making the participant feel uncomfortable. It is possible to support this experience with sound, light and a suitable input device.

The different input devices that are available in the VRLab and could be suitable for this research are firstly a simple Tablet or Smartphone, which can be attached to the Theatre Screen which shows the user interface of the artificial pancreas and can be operated via touch of the buttons; or a mockup model of the artificial pancreas. A simplified model of the product could be built by using a simple screen technique and buttons that feel similar to the ones in the end product and putting it in a model of the same size and similar weight of the artificial pancreas. If Inreda BV has the possibility to build a mockup model it could be attached as an input device and would give the participants a very realistic feel of the device.

To conclude the selection of methods for virtual reality methods, it can be summarized that VR in general is a good solution for this research and the most suitable techniques are laboratory testing where the co-creator interacts with a virtual environment projected on a Theatre Screen with a mockup model of the artificial pancreas or if not available a Tablet or Smartphone that shows the user interface. The mockup model would be slightly better, because it would also allow the user to push the actual buttons and hold the device, but both should lead to a good outcome of the study.

2.5 Application in the case of Inreda BV

As introduced before, Inreda BV is the developing company of the artificial pancreas of which the usability should be tested according to how this research proposes.

The artificial pancreas is until now designed for the use of adult diabetes 1 patients and therefore also in the testing all co-creators will be adults. The use of such a device for children would create new challenges, which will be coped with in the future.

Inreda BV has just designed a new model for an artificial pancreas that is smaller (90 x 76,5 x 21-25 mm) than the latest prototype and has a new user interface that is supposed to be easier to operate (Inreda Diabetic BV, 2015).

The developing company of the artificial pancreas is especially keen on knowing if the newly designed interface of the device is easy to use and if it includes all information that is needed by the patient to ensure a reliable control of glucose levels and the condition of the device. The patient needs to be able to monitor his or her own health status as well as the battery level or level of insulin and glucagon cartridges easily. He or she also needs to be able to change cartridges without problems (Inreda Diabetic BV, 2015).

Next to the user interface, Inreda BV would like to find out how patients react to alarms that go off if something is wrong with the patient's status or the status of the device. It is interesting to see whether it is useful to have alarms that go off at night, and if so, which alarms are important enough to disturb the patients sleep and which alarms can wait until the next morning.

To test those specific situations and the reaction of patients, the following design approach is proposed.

3. RECOMMENDATIONS AND DESIGN APPROACH

Taking all aspects tested above into account, the usability test of an artificial pancreas in a virtual environment involving diabetes 1 patients as co-creators could look as follows.

3.1 Selection of Participants

The selection of the right participants is crucial for a reliable and valid outcome of the test. As explained earlier patients with all different kinds of backgrounds are needed in order to ensure that the outcome is generalizable for adult diabetes 1 patients. The group of patients in the study should include different age groups, individuals with different levels of technical skills, patients with different life-styles (e.g. patients that are involved in sports, patients that work desk-jobs and do not work out in their free time), patients with different levels of body fitness and size, etc. Those characteristics can be determined by a preliminary questionnaire that should be sent out to a large group of diabetes 1 patients who are willing to participate in the study. (See Appendix A) Inreda BV keeps a list of around 3000 diabetes type 1 patients that are willing to bring diabetes treatment methods forward and are therefore motivated to participate in studies. The preliminary questionnaire could be send to all patients on this list to then receive enough responses in order to find suitable participants.

According to the returned questionnaires the developing company can select a group of patients that cover as many different characteristics as possible (e.g. level of body fitness, eating habits, technical skills, age, gender, etc.).

3.2 Test arrangement and tasks

The test arrangement in the VRLab of the University of Twente will look as follows: a Tablet or a mockup artificial pancreas is connected to the Theatre Screen. On the screen of the Theatre Wall, the user will see an avatar that acts in a virtual world in a first-person setting. On the device he or she will see the interface of the artificial pancreas, as it would look in the finished product. The avatar will start to walk through different every-day situations that are likely to happen (sometimes unexpected) and the co-creator will have to operate the device if necessary to ensure a healthy glucose level, as if he would move through the virtual environment himself.

The different situations could look as follows:

3.2.1

On the screen appears a clock that signals that it is morning and the avatar leaves the house to catch the bus. As he/she walks on the street he/she walks on the street he/she sees the bus standing at the bus stop and starts running, which perfectly fits to the stressful morning. When he/she finally sits on the bus it is the first calm moment and the avatar decides to check its glucose levels for the past hours. On the top of the screen a task appears for the co-creator: Please check your current blood glucose level and the trend for the last 10 hours.

3.2.2

The avatar is now sitting at a desk (at his/her job) when a signal starts and the right part of the devices monitor, which signals the status of the device, turns orange. The level of the glucagon cartridge is low. The co-creator sees a task on the top of the screen: Please check, if the glucagon cartridge is full enough to bring you through the day.

The patient needs to select the right buttons to check the estimated duration of the cartridge and evaluate if he/she needs to change the cartridge or if it has enough glucagon for the day until he/she gets home.

The status on the device should show a sufficient time so that the right solution to the task is: Don't change cartridge yet.

3.2.3

In the third scene the avatar gets home and a red alarm appears on the right side of the device monitor and a loud warning signal appears. The user sees a task on the top of the screen: Please change the glucagon cartridge according to the manual on the device's screen.

The device automatically shows a manual that helps the user to change the cartridge step by step, therefore the co-creator is given no additional information. After every step the user needs to confirm that the step is done on the device's interface. The change of the cartridges appears in the picture when the user selects the "Change cartridge" button and will be manipulated after the user finds the right button on the device. To ensure that the handling of the input device does not affect the efficiency and effectiveness measurements, the cartridges slip into place on the screen and the user only needs to select the right buttons.

3.2.4

The next scene plays during the night where the patient peacefully sleeps in his/her bed. The alarm clock on the nightstand shows 04:17 o'clock. Suddenly an unexpected alarm goes off. The battery of the device is low. With slower movements and worse vision than usual, (according to the abilities of a person that just woke up due to an alarm) the co-creator needs to identify the alarm and change or charge the battery as shown on the display of the device.

3.2.5

In the evening the avatar has dinner with friends. As this meal is different from his/her usual eating habits more insulin is needed. A task appears on the top of the screen: Please check your blood glucose level manually (with a finger prick) to check whether the sensors are working fine, according to the instructions on screen.

The task is completed once the user recognizes that the avatar pricked his/her finger and put the test strip into the device and the user presses the right button to calibrate the data.

3.2.6

Apparently the avatar has had many dinners with friends lately and sees that he/she has gained 5 kg when he/she looks at the scale the next morning. Since the device takes the body weight into account when it calculates the right amounts of insulin or glucagon the body weight entered in the device always needs to be updated. The next task appears on the top of the screen: Please adjust your body weight on the artificial pancreas.

The task is completed when the user has entered the new body weight and pressed the confirm button.

3.3 Effectiveness and Efficiency testing

As explained in 3.2.1. and 3.2.2. the effectiveness and efficiency, as parts of the usability of the device, are tested with the observation through experts. The effectiveness of the artificial pancreas is tested by counting the number of tries a patient needs to complete a task. The efficiency is tested by taking the time a patient needs to complete a task.

This quantitative data conducted by the experts needs to be measured with the same methods and circumstances for all patients participating in the experiment to make it comparable with one another and ensure validity of the outcome of the study.

3.4 Post-Test Questionnaire

After the virtual reality test a direct questionnaire should be answered by the co-creators. This questionnaire will deliver the data that can later be transformed into a measure of satisfaction. The questionnaire should hold detailed questions about how the patient felt in different situations during the test, how much effort was needed to figure out how to complete the task or understand alarms on the device, which buttons or information might be missing on the user interface according to the different co-creators, or the overall satisfaction of the user with the artificial pancreas.

The style of the questionnaire should consist of different kinds of answer types. Questions can be answered according to a scale, multiple choice or open questions. (See Appendix B)

3.5 Analyzing Satisfaction

The satisfaction of co-creators with the device can be analyzed using the direct post-test questionnaire. Qualitative data can be transferred to quantitative data and then compared with the data of all co-creators. As an addition, comments of the respondents should be taken into account to receive extra information about individual thoughts and feelings. With the extra comments, ideas about latent needs or implications for improvements can be communicated and analyzed. Latent needs can also be detected during the conduction of the test, if the co-creators are asked to speak out loud during the test. Manon Spin and Roy Damgrave (Appendix D and E) both suggested that the co-creators should tell what they are doing and what they are thinking. The experts that are observing the test should collect any indications or reactions to get an overview of what the person conducting the test was thinking and feeling at the moment he/she was confronted with a specific situation or task.

3.6 Comparison and statistics

In the end of the study, all data should be compared and ratios for effectiveness, efficiency and satisfaction can be created to determine the overall usability of the artificial pancreas.

4. CONCLUSION

To conclude the findings of this research paper one can say that the usability of an artificial pancreas can be tested in a virtual reality environment involving the end-users as co-creators according to the design approach that has been proposed in this paper. This design approach that involves direct questionnaires and the observation through experts should lead to comparable data that can bring the desired value to the developing company in terms of information on the usability of the artificial pancreas, patient's wishes, wants or latent needs, and value in monetary terms due to savings in the test phase compared to trial and error prototype testing.

The design approach for a test in the VRLab of the University of Twente can be seen as the answer to the research question of this paper.

Inreda BV should be able to use this test approach and test the usability of their artificial pancreas, which brings the innovation of diabetes type 1 treatments a little step forward. It would also be possible to take the design approach and make small changes to it in order to use it for the test of other generations of models later on.

Even though it was possible to come to an answer of the research question, there were several limitations to this research. Firstly the time frame for this research was limited to a total of 9 weeks, because it was written in the context of a bachelor thesis, which might have an influence on the depth of the theory. Furthermore the length of this research paper should not extent the predefined number of pages, which also put a limitation on the depth and elaborateness of the research paper. Another limitation that is also linked to the other two limitations, though, was that it was not possible to conduct the research within the frame of this research paper. The outcomes of the designed test approach are therefore open, but leave an opportunity for further research studies (e.g. another Bachelor or Master Thesis).

More implications for further research are in how far the usability of the artificial pancreas is connected to the perceived value-in-use and how this could be tested in a VR environment, as well as how an artificial pancreas could be used in the case of children that suffer from diabetes 1. Furthermore, I had to limit my study on the usability of the user interface and the alarms, which are both related to the software of the product, while it would also be necessary to test the usability of the hardware of the artificial pancreas, as Roy Damgrave describes in his interview, the hardware could influence the outcomes of the software testing (See Appendix E) and Manon Spin also proposes the hardware testing as an implication for further research (Appendix D). Additionally it would be interesting to further use the usability dimensions by Han et al. (2000), for a more detailed usability test of the software as well as the hardware of the device.

5. ACKNOWLEDGEMENTS

I would like to express my very great appreciation to my first supervisor Tamara Oukes and my second supervisor Ariane van Reasfeldmeijer for taking the time and patience to supervise my bachelor thesis and always providing me with valuable feedback, comments or ideas. The assistance given by Manon Spin and Roy Damgrave who were so kind to answer my questions about the viability of my test design in an interview was greatly appreciated. Furthermore I would like to thank my family and friends for their support and encouragement throughout my study.

6. REFERENCES

Anderson, J. C., & Narus, J. A. (2004). Market sensing. In Business market management: Understanding, creating, and delivering value.

Babbar, S., Behara, R., & White, E. (2002). Mapping product usability. International Journal of Operations & Production Management. doi:10.1108/01443570210446315

Babbie, E. (2010). The practice of social research (13th ed.). Wadsworth.

Blach, R., Landauer, J., Rösch, A., & Simon, A. (1998). A highly flexible virtual reality system. Future Generation Computer Systems. doi:10.1016/S0167-739X(98)00019-3

Booz&co. (2011). Diabetes Care in the Netherlands: Improving health and wealth.

Business Insights Ltd. (2011). The Diabetes Device Market Outlook to 2016. Retrieved from BI00043-001

Carrozzo, M., & Lacquaniti, F. (1998). Virtual reality: a tutorial. Electroencephalography and Clinical Neurophysiology/electromyography and Motor Control, 1-9. doi:10.1016/S0924-980X(97)00086-6

Cruz-Neira, C. (1998). Making virtual reality useful: A report on immersive applications at Iowa State University. Future Generation Computer Systems. doi:10.1016/S0167-739X(98)00017-X

Di Gironimo, G., Matrone, G., Tarallo, A., Trotta, M., & Lanzotti, A. (2013). A virtual reality approach for usability assessment: case study on a wheelchair-mounted robot manipulator. Engineering with Computers, 359-373. doi:10.1007/s00366-012-0274-x

Eliashberg, J., Lilien, G. and Rao, V. (1997). Minimizing Technological Oversights: A Marketing Research Perspective. Technological Oversights and Foresights Conference, New York.

Füller, J., & Matzler, K. (2007). Virtual product experience and customer participation—A chance for customer-centred, really new products. Technovation. doi:10.1016/j.technovation.2006.09.005

Guttentag, D. A. (2010). Virtual reality: Applications and implications for tourism. Tourism Management. doi:10.1016/j.tourman.2009.07.003

Han, S. H., Yun, M. H., Kim, K., & Kwahk, J. (2000). Evaluation of product usability: development and validation of usability dimensions and design elements based on empirical models. International Journal of Industrial Ergonomics. doi:10.1016/S0169-8141(00)00019-6

Inreda Diabetic BV. (2015). New Model. Retrieved October 25, 2015, from www.inredadiabetic.nl/product/new-model/

International Diabetes Federation. (2013). IDF Diabetes Atlas (6th ed.). Retrieved from ISBN: 2-930229-85-3

Klaver, L. (2015). A review of research techniques for co-creation and an investigation into patient needs and product requirements to study the intention to use of a new medical device to regulate diabetes 1 (Master's thesis, University of Twente, Enschede, The Netherlands). Retrieved from http://essay.utwente.nl/67750/

Kuhlen, T., & Dohle, C. (1995). Virtual reality for physically disabled people. Computers in Biology and Medicine. doi:10.1016/0010-4825(94)00039-S

Kujala, S. (2003). User involvement: a review of the benefits and challenges. Behaviour & Information Technology, 22(1), 1-16. doi:10.1080/01449290301782

Leminen, S., Westerlund, M., & Nyström, A. G. (2012). Living labs as open-innovation networks. Technology Innovation Management Review, 6-11.

McColl-Kennedy, J. R., Vargo, S. L., Dagger, T. S., & Sweeney, J. C. (2009). Customers as resource integrators: Styles of customer co-creation. ResearchGate.

Pan, Z., Cheok, A. D., Yang, H., Zhu, J., & Shi, J. (2006). Virtual Reality and Mixed Reality for Virtual Learning Environments. Computers & Graphics. doi:10.1007/11736639_4

Prahalad, C. K., & Ramaswamy, V. (2004). Co-creation experiences: The next practice in value creation. Journal of Interactive Marketing. doi:10.1002/dir.20015

Rizzo, A., & Kim, G. J. (2005). A SWOT Analysis of the Field of Virtual Reality Rehabilitation and Therapy. Teleoperators and Virtual Environments, 14(2), 119-146. doi:10.1162/1054746053967094

Suh, K., & Lee, Y. E. (2005). The Effects of Virtual Reality on Consumer Learning: An Empirical Investigation. Management Information Systems Quarterly.

University of Twente. (2012). VR Laboratory - University of Twente. Retrieved August 20, 2015, from http://www.vrlab.ctw.utwente.nl/

Van Bon, A. C., Hermanides, J., Knoops, R., Hoekstra, J. B., & DeVries, J. H. (2010). Postprandial Glycemic Excursions with the use of a Closed-Loop Platform in Subjects with Type 1 Diabetes: A Pilot Study. Journal of Diabetes Science and Technology, 4(4), 923-928. Retrieved from www.journalofdst.org

Van Bon, A. C., Jonker, L. D., Koebrugge, R., Knoops, R., Hoekstra, J. B., & DeVries, J. H. (2012). Feasibility of a bihormonal Closed-Loop System to control postexercise and postprandial Glucose Excursions. Journal of Diabetes Science and Technology, 6(5), 1114-1122. Retrieved from www.journalofdst.org

Van Bon, A. C., Luijf, Y. M., Koebrugge, R., Knoops, R., Hoekstra, J. B., & DeVries, J. H. (2014). Feasibility of a portable bihormonal Closed-Loop System to control Glucose Excursions at Home under free-living Conditions for 48 Hours. Diabetes Technology & Therapeutics, 16(3), 1-6. doi:10.1089/dia.2013.0166

Van Kleef, E., Van Trijp, H. C., & Luning, P. (2005). Consumer research in the early stages of new product development: a critical review of methods and techniques. Food Quality and Preference, 16(3), 181-201. doi:10.1016/j.foodqual.2004.05.012

7. APPENDIX

A. Questionnaire for selection of participants

Please fill in the following questionnaire to determine if you are a suitable participant for our usability test. You will be informed about the selection soon after sending in the questionnaire.

Do you suffer from Diabetes 1?

- ○ Yes
- ○ No

For how long do you suffer from Diabetes 1?
Please write down the number of years.

[ ]

How old are you?

[ ]

What is your current height?
in cm

[ ]

What is your current weight?
in kg

[ ]

How many hours are you physically active per week?
Please select the answer that fits best. (Physical activities include walking, cycling to work, cleaning the house, walking the dog, etc.)

- ○ < 3 hours
- ○ 3 - 5 hours
- ○ 5 - 7 hours
- ○ > 7 hours

As how physically challenging would you describe your job?
Please evaluate how much physical activity your job requires on a scale from 1 to 5.

	1	2	3	4	5	
Desk job (not very active)	○	○	○	○	○	Highly physical job

How many hours do you physically exercise per week?
Please select the answer that fits best. (Physical exercise means jogging, swimming, team sports, tennis, yoga, etc.)

- < 3 hours
- 3 - 5 hours
- 5 - 7 hours
- > 7 hours

How many times do you eat meals or snacks per day?
Please fill in the answer according to an average day.

- < 3 times
- 3 times
- 4 times
- 5 times
- > 5 times

As how sugar-containing would you describe your eating habits?
Please evaluate how sugar- containing your eating habits are on a scale from 1 to 5.

	1	2	3	4	5	
Very sugar-containing	○	○	○	○	○	Little sugar-containing

How familiar would you describe yourself with the latest technique?
Please evaluate how familiar you are on a scale from 1 to 5.

	1	2	3	4	5	
Not at all	○	○	○	○	○	Very familiar / heavy user

In which group would you put yourself when it comes to new technique on the market?
Please select the group you can identify yourself with the best.

- Innovator (being the first to try new ideas, goods or services)
- Early adopter (being very early to try new ideas, goods or services)
- Early majority (wanting to try new ideas, goods or services as soon as they are proven useful by the early adopters)
- Late majority (adopting innovations only if you're being forced to and are not swayed by advertisement)
- Laggard (you dislike change and stick with products or services you already have as long as possible)

Would you be willing to participate in the usability testing of an artificial pancreas in order to bring innovation in diabetes care forward?

- Yes
- No

B. Post-Test Questionnaire

Please fill in the questionnaire below after participating in the virtual reality test of the artificial pancreas.

How comfortable did you feel during the experiment?
Please give your answer on a scale from 1 to 5.

	1	2	3	4	5	
Very uncomfortable	○	○	○	○	○	Very comfortable

In the first scene, how did you feel when you wanted to check your current glucose level and the trend for the last 10 hours?
Please select the adjectives that fit best with your feelings.

- ○ Relieved
- ○ Insecure
- ○ Secure
- ○ Stressed
- ○ Vulnerable
- ○ None of the above
- ○ Other: [ ]

How much effort did it take to complete the task?
Please give your answer on a scale from 1 to 5.

	1	2	3	4	5	
Not much / easy	○	○	○	○	○	A lot / difficult to find

Do you have suggestions for improvement related to this situation?
Please feel free to name anything that you would change.

[ ]

In the second scene, how did you feel when the alarm signaled a low level of glucagon in the cartridge?
Please select the adjectives that fit best with your feelings.

- ○ Nervous
- ○ Insecure
- ○ Secure
- ○ Stressed
- ○ Vulnerable
- ○ Calm
- ○ None of the above
- ○ Other: [ ]

How much effort did it take to complete the task?
Please give your answer on a scale from 1 to 5.

	1	2	3	4	5	
Not much / easy	○	○	○	○	○	A lot / very hard

Do you have suggestions for improvement related to this task?
Please feel free to name anything that you would change.

In the third scene a red signal and loud alarm appeared, how did you feel in this situation?
Please select the adjectives that best describe your feelings.

- ○ Nervous
- ○ Insecure
- ○ Scared
- ○ Secure
- ○ Stressed
- ○ Vulnerable
- ○ Calm
- ○ None of the above
- ○ Other:

How much effort did it take to find out how you can solve the problem?
Please evaluate your answer on a scale from 1 to 5.

	1	2	3	4	5	
Not much / easy to solve	○	○	○	○	○	A lot / very unclear

Do you have suggestions for improvement related to this alarm?
Please feel free to name anything that you would change.

How much effort did it take to complete the task and change the cartridge?
Please evaluate your answer on a scale from 1 to 5.

	1	2	3	4	5	
Not much / easy to complete	○	○	○	○	○	A lot / very hard to complete

Were the instructions on the device clear?

- ○ Yes
- ○ After some consideration
- ○ No

Do you have suggestions for improvement related to this task?
Please feel free to name anything that you would change.

In the fourth scene, you were suddenly woken up by an alarm, how did you feel when the alarm started?
Please select the adjectives that best describe your feelings in the situation.

- Confused
- Angry
- Scared
- Insecure
- Calm
- Secure
- Nervous
- Vulnerable
- None of the above
- Other:

How much effort did it take to find out what caused the alarm?
Please evaluate your answer on scale from 1 to 5.

	1	2	3	4	5	
Not much / easy to determine						A lot / unclear and confusing

How much effort did it take to find out how to solve the problem and recharge the battery?
Please evaluate your answer on a scale from 1 to 5.

	1	2	3	4	5	
Not much / easy to find out						A lot / unclear and confusing

Were the instructions on the device clear?

- Yes
- After some consideration
- No

Do you have suggestions for improvement related to this task?
Please feel free to name anything that you would change.

In the next scene, your task was to view and calibrate your glucose level during dinner. How much effort did it take to find this function?
Please evaluate your answer on a scale from 1 to 5.

	1	2	3	4	5	
Not much / easy access to data						A lot / hard to find data

How did you feel in this situation?
Please select the adjectives that best describe your feelings in this situation.

- ○ Confused
- ○ Insecure
- ○ Helpless
- ○ Calm
- ○ Secure
- ○ Scared
- ○ Nervous
- ○ Vulnerable
- ○ None of the above
- ○ Other: [ ]

Do you have suggestions for improvement related to this task?
Please feel free to name anything that you would change.

[ ]

In the next scene you were asked to change your body weight on the device. How much effort did it take to figure out how to complete the task?
Please evaluate your answer on a scale from 1 to 5.

	1	2	3	4	5	
Not much / easy access to data	○	○	○	○	○	A lot / hard to find and change data

Do you have suggestions for improvement related to this task?
Please feel free to name anything that you would change.

[ ]

Overall, how did you feel when operating the device?
Please evaluate your answer on a scale from 1 (insecure) to 5 (secure).

	1	2	3	4	5	
Insecure	○	○	○	○	○	Secure

How satisfied are you with the user interface of the device?
Please evaluate your answer on a scale from 1 to 5.

	1	2	3	4	5	
Not satisfied	○	○	○	○	○	Very satisfied

Do you have suggestions for improvement related to the user interface?
Please feel free to name anything that you would change.

[ ]

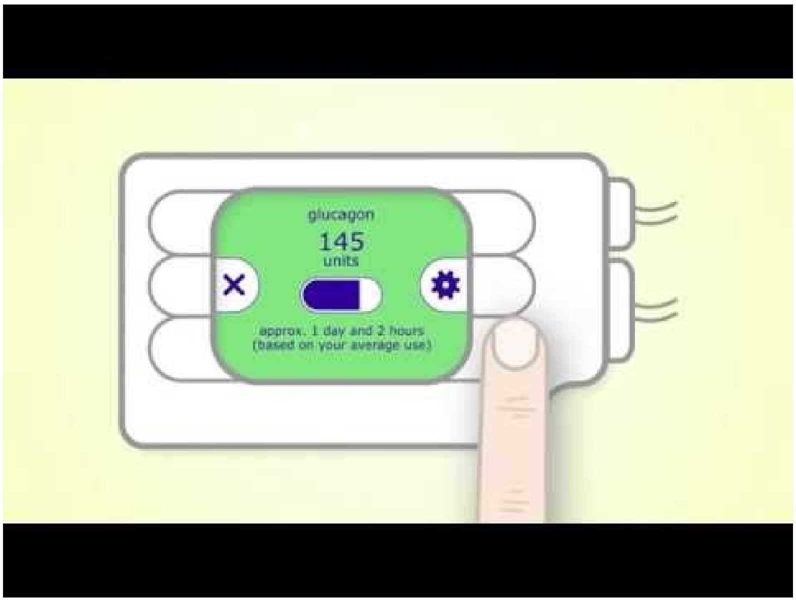
How satisfied are you with the alarms that occur if something is wrong with you or the device?
Please evaluate your answer on a scale from 1 to 5.

	1	2	3	4	5	
Not satisfied	○	○	○	○	○	Very satisfied

Do you have suggestions for improvement related to the alarms of the device?
Please feel free to name anything that you would change.

Please feel free to give any remarks you have on the artificial pancreas after testing it in the virtual reality laboratory.

C. The new model's interface



D. Interview with Manon Spin

Interview with Manon Spin (expert in social research) 28.10.2015

1) How many participants would be needed to send out the selection Questionnaires?

Definite response rates can be found for different kinds of tests or experiments. According to my knowledge Inreda BV keeps a list of possible participants who are motivated to bring diabetes care innovation forward.

2) How many participants need to be selected to make the outcomes of the study valid?

It depends on the kind of test and desired outcome. In this case, I would propose to use an approach where a smaller group (for example 10) of participants is selected and conducts the test. After that, the same test is conducted with a second group (again 10 people). If you can see a difference in answers in the second group, try a third group. This should be done until no new information is collected with the Post-Test questionnaires.

3) Are the types of questions suitable for this study?

Yes, some Questions need to be formulated sharper to ensure that every participant interprets it right, but the type in general is suitable.

4) Do I need more open or more closed questions?

The mix of open and closed questions seems to be ok.

5) Are scales suitable to bring me to the desired outcome?

In some cases they are suitable, but in other cases more definite answer possibilities (like multiple choice) are more suitable. It is for example difficult to ask a person “how active he/she is in general” because people might have a different feeling about what activity is. It would be better to ask a person for “the amount of hours spent of physical activity per day/week”, including examples like walking, cycling to work, walking the dog etc.

6) Is the plan of selection Q., first impression Q., and Post-Test Q., suitable and will bring a good outcome?

In general yes, but since the first impression questionnaire mainly focuses on the hardware of the device and your study is more focused on the software (user interface and alarms) I would suggest to leave it out and maybe take it as an implication for further research.

7) What do you think of the efficiency and effectiveness measures?

In combination with the satisfaction measure it is a good way to test the usability, because it combines direct observation and indirect feedback of the participant. The questionnaires alone would not be fully reliable, because we can never be 100% sure that the answers are reliable.

8) Do you think I need to include anything else in the tests?

I would suggest to use a method in addition that allows collecting a more reliable feedback of the feelings or emotions of a participant in the very moment of the situation, in which the participant is asked to “talk out loud” during the test. The experts that are present for the effectiveness and efficiency test could then take notes about how the participant reacts.

E. Interview with Roy Damgrave (VRLab University of Twente)

Interview with Roy Damgrave **29.10.2015**

The situation: I’m writing my thesis about how the usability of an artificial pancreas can be tested in a virtual reality laboratory involving end-users as co-creators. I evaluated methods in my thesis to figure out which testing methods are most applicable. In the end I came up with a design approach for the test. I would now like to know whether this is possible in the VRLab or not.

I already checked the available technique in the VRLab and evaluated which devices are most applicable for this test.

The test will be conducted with one patient at a time.

Scenes of the daily life will be shown on an Elumens to make it look most realistic to the patient. On screen an Avatar will go through the day and on the bottom of the screen the user interface of the artificial pancreas will be shown to enable the patient to interact with the interface.

1) Are Elumens and a Phantom Omni useful devices for the desired outcome?
It is not the ideal solution, because Elumens are not the latest technology anymore and participants of the test are sitting on a “desk”, which does not completely deliver the feeling you want to bring across. The screen is also relatively small, so that you still have the surroundings in the lab visible, which doesn’t necessarily have to be a problem, but it needs to be taken into account when choosing the equipment for the test.
The Phantom Omni is basically a haptic device in the shape of a pen. Due to the fact that the user is constantly holding the pen in his/her hand, the freedom of movement is limited and since you want to create the feeling of touching the buttons of the user interface of the artificial pancreas, it is not the best possibility.

2) Can the Scenes be realized on the chosen screens?
The scenes can be realized on pretty much any screen in our laboratory. Since the surrounding on the screen is only supposed to give the participant an idea of the situation, but is not necessary for the interaction I would suggest that also a photo of the surrounding shown an appropriate screen could be sufficient. Depending on the screen technique this picture can be shown on a large screen or even a 360° degree picture in which the participant can move around with an oculus rift.

3) Do you have recommendations for more applicable screen technique?
The oculus rift could perhaps be a good solution, because it allows the user to fully emerge in the virtual environment and allows direct interaction with the users’ own hands in the virtual environment. It might

be a little bit too much for some participants, though, who are not familiar with virtual reality and cause uncomfortable feelings and different reactions.
Another possibility would be to use the big theatre screen, so that the participant can still see the real environment around him/her if he/she wants to. The size of the theatre screen allows the user to also focus on the screen and the virtual environment without looking at the real environment.
The picture on the theatre screen comes from a projector that is fixed on the wall behind the participant, though, which sometimes can lead to unwanted shadows on the screen that disturb the experience.
Therefore the touch wall would also be a possible solution, on which the picture comes from projectors behind the glass. The picture is therefore not disturbed by the participant. The touch wall is a little bit smaller than the theatre screen, though.

4) How much technical skill is needed to use a Phantom Omni?
The Phantom Omni is, as we discussed before, not the best option in this study. For the devices we discussed as suitable (Oculus Rift, Tablets or Smartphones, a simple model of the artificial pancreas that is developed to show the user the real interface and buttons) not much technical skill is needed to operate it. With the oculus rift, interaction is possible through simple hand movement and on tablets or smartphones the participant would see the real device interface. Through the simple touch screen, not much technical skill is needed.

5) Do you have recommendations for input devices that are easier to manipulate?
The Oculus Rift does not need additional input devices, but could be added an additional model of the artificial pancreas on the participant's waist to make the experience more realistic. The participant could then look down and manipulate the buttons on the model device with his/her own hands and would see the reaction of the device virtually.
A tablet or smartphone could act as an input device when using the theatre screen or touch wall for displaying the virtual world. The touch screen of the input device would show the user interface of the artificial pancreas and allow the participant to select buttons on the screen.
My advice would be to ask Inreda BV to create a simple mockup model of the artificial pancreas that allows the user to press the actual buttons and look at the actual screen of the device. The participant could then also experience to carry the mockup device around the waist and feel its size and weight.

6) Do you have any additional ideas or recommendations?
It would be a good idea to include sound as an extra feature to create a deeper feeling of the situation. In the first scene when a bus is shown, the actual sound of a crowded bus can help the virtual reality experience.
This would also help to better embed the alarm that is tested.
The mockup model of an artificial pancreas seems like a very suitable idea to me, which would combine virtual reality and real world experience.
You have to be aware of what you want to test, though. The way of interaction (display with 6 buttons) is already chosen; do you want to test the GUI on this screen, or the position of the buttons, or the shape of the device/screen, etc.. What is the goal of the test and what should it bring at the end? Is that an advice for redesign or a good/not good decision?
And what is allowed to change to the interaction; so what are the variables you can influence after the test.
You have the risk that you are going to test your own boundaries: the design decision already made during the product development.