

Development and trial of an online survey to investigate trust before the use of a technology

Abstract

Trust is an important concept in social interaction. Research shows that trust is also of importance when people interact with technology. Extensive research has been conducted to measure trust when using technology. No research so far has looked into trust before the use of a product, which is of great importance for new products on the market. The current work aims at developing a tool (survey) to measure trust in medical devices for home usage (Home MDD's) and testing the usability of that tool. The survey has been constructed in tune with the literature. Experts were asked to make a list of Home MDD's and associated features that have been used in the survey. Participants filled the survey and were interviewed using a Retrospective Thinking Aloud (RTA) method. Suggestions for redesign were made based on the data of the interviews.

1. Introduction

“no mortal can keep a secret. If his lips are silent, he chatters with his fingertips; betrayal oozes out of him at every pore”. Sigmund Freud (1905) talked about how people judge each other and how the author believes one cannot trust another person. Studies that followed showed how people can control some of their communication channels, but not all (Ekman & Friesen, 1974). Some communication channels (i.e. posture, facial expressions) can show uncontrolled messages. People have been used to judging part of their trust, or distrust, in another based on visual cues (Bond Jr & DePaulo, 2006, 2008; Vrij, 2008). Trust in its broad sense could be defined as follows: “*Trust is the willingness of the trustor to rely on a trustee to do what is promised in a given context, irrespective of the ability to monitor or control the trustee, and even though negative consequences may occur.*” (Aljazzaf, Perry, & Capretz, 2010). This definition shows that trust is an interaction between at least two parties, the trustor and the trustee, and is usually applied in human to human exchanges. Following (Kelton,

Fleischmann, & Wallace, 2008), trust has four different levels that could be summarized as follows: i) individual e.g., I trust, ii) the interpersonal e.g., I trust you, iii) the relational e.g., You and I trust each other, and iv) the societal level e.g., We all trust.

In a society that is increasingly making use of technologies to support daily and corporate life, a new forms of interaction is becoming of bigger importance each day, namely human-technology interaction. In human-technology interaction, technology substitutes for one of the human actors in the interaction. Applied to the concept of trust, technology would substitute for either the trustor or the trustee. But can technology, being designed by humans, act as either a trustor or a trustee? It would be naive to say that one can trust the person who designed the technology instead of the technology itself. Different studies about trust in vendors on the internet support this view (Gefen, Karahanna, & Straub, 2003; Kim, 2008; Lim, Sia, Lee, & Benbasat, 2006). In these studies, the uncertainty of the internet plays a significant role. This uncertainty is a result of characteristics the internet possesses, like being prone to alteration (Flanagin & Metzger, 2000, 2007) and a lack of standardized methods and behavioral cues that serve to increase the development in online trust (Kelton et al., 2008; Rocco, 1998). Another study that lays focus on the process of trust in another person through a digital environment shows that communication over the internet may increase the perception of vulnerability and therefore decrease trust (Rocco, 1998). These are examples of human-human interaction, however, we are also talking about characteristics of the internet as catalyzers for the lack of trust (i.e. being prone to alteration). So is it possible to have a feeling of trust towards these characteristics?

The possibility for trust in technology itself, that is, directly trusting the technology instead of the person or company behind that technology, has been a subject for discussion in past literature. Some researchers have denied that people can trust technology at all. Friedman and colleges (2000), in line with other researchers (Lynch, 2001; Solomon, 2000) concluded that *"people trust people, not technology"*. An increasing amount of researchers,

however, is starting to accept that it is also possible to trust technology itself (Lankton, McKnight, & Tripp, 2015).

Lankton and McKnight (2011) stated that only a small amount of literature involves trusting technology itself, where technology is seen as an IT artifact instead of a person or company. The research conducted by McKnight, Choudhury, and Kacmar (2002) aimed for an understanding of the difference between trust in people and trust in technology. Their theoretical framework consists out of known attributes related to trust in human-human interaction and their technological counterpart. Specifically mentioned are 'competence', 'benevolence' and 'predictability'. Their technological counterparts are supposed to be 'functionality', 'helpfulness' and 'reliability' respectively. What follows from this comparison is that, in order to trust technology, certain conditions have to be met that are comparable with positive conditions for trust in people. These conditions are, however, not exactly the same. Therefore, a distinction can be made between trust conditions that are related to human characteristics and conditions that are related to specific characteristics of technology. The former will be named 'human-related conditions' and the latter will be named 'technology-specific conditions' for the rest of this paper.

For interaction with technology, technology-specific conditions are not always the correct way to measure the level of trust in that technology. As an example, Lankton and McKnight (2011) found that a group of Facebook users trust Facebook not only as a technology, but also trust Facebook as a quasi-human. Building on this, Lankton et al. (2015) advise researchers to use multiple conditional factors to assess trust in technology. There are two different trust constructs, namely human-like and a technology-like. It is advocated that the human-like trust construct is applicable to technologies that have human-like features (a voice or some animated components) while the technology-like trust construct is mostly applicable to systems that do not have these features. An example of a technology that has human-like features is Apple's Siri, which has a voice functionality. A voice functionality represents the

human characteristic of talking and being able to make conversation, which confirms to the human-like trust construct. An example of a technology that does not have these human-like features is Microsoft Excel, which is basically a 'simple' spreadsheet and represents only characteristics of technology. In order to measure trust in technology one has to understand the degree of humanness of that technology and use the relevant attributes (Lankton et al., 2015).

Most research on trust in technology has been conducted after the technology was used at least one time, where the trust in that technology comes forth out of the use case (Lucassen, 2013). There are, however, indications that establishing trust in technology before the use is mainly driven by visible cues and perceived quality of the product. A well-established paradigm to support this is the emotional design of Norman (2004) which states that the emotional design of a product may be far more critical to the success of a product than the practical design (Fishwick, 2004). Following this reasoning, a product that has a beautiful design makes people feel good, and in turn this good feeling makes people more willing to be creative and find solutions to the problems they encounter while using the product. Norman (2004) differentiates three levels of design.

- The first level is the 'visceral' level, covering the sensory aspects of a certain product (look, feel, smell, sound).
- The second level is the 'behavioral' level, where users form an opinion about the product based on a first use case.
- The third level is the 'reflective' level, containing the sentiment that arises after the product has been used for some time.

The first level of this paradigm is important for the current research into trust established by aesthetics.

1.2. Aim of the present work

The aim of this work is to develop a digital tool (survey) to enable the assessment of trust

towards that product before the use. Trust can be measured using multiple strategies, ranging from trust games (Buchan, Croson, & Solnick, 2008) to survey's (Naef & Schupp, 2009). The use of trust games usually requires multiple human participants. Because the aim of the current work, measuring trust in technology, a choice has been made for a survey measurement. A survey only requires a single participant and can measure trust toward for example a technology. Next to designing the tool, the current work will assess the usability of this tool to inform future redesign. Specifically this study will focus on a tool for measuring trust in medical devices for home usage (Home MDD's)

We are not aware of previous studies to investigate the change in trust, associated with an increasing amount of information, before the use of a new product using an online digital tool. In particular, we will focus the present survey on the trust before the use of medical devices for home usage. To achieve the aim, three main phases are performed.

- i. *Design of the survey*: Starting from a paper prototype of the survey developed by Dr. Borsci in tune with literature and previous studies (See: Appendix A), we developed a digital tool to be usable online and to enable the safe gathering of participant's opinions about their trust towards four blood pressure monitors (BPMs). The initial set of BPMs (and its associated list of features) was retrieved from internet. Four devices were selected as potential candidates for the survey by the researcher of the present work and Dr. Borsci.
- ii. *Expert agreement BPMs and definition of the stimuli*: Using the description of the four devices we created a list of 24 features (Appendix B) used by manufacturers to described the BPMs to potential consumers. Experts were asked to categorized these 24 features in terms of usability and aesthetics.
- iii. *Usability evaluation of the digital survey*: We tested the functionalities of the survey in terms of data export. Then we involved a group of participants to use and to evaluate the usability of the digital survey.

1.3. Why trust is important

With the increasing use of technology by consumers, companies and governments there is an increasing need for trust in those technologies and services. Trust is a crucial factor for the success of companies, technologies and online systems (Hansen, Saridakis, & Benson, 2018). It is shown that organizations lacking a profound reputation are seen as less trustable (Jarvenpaa, Tractinsky, & Vitale, 2000). The same can be said for new and upcoming technologies that have not yet withstood the test of time.

For the time being, most research has focused on measuring trust after the use of a product, however, new products that have a good quality will not instantly be bought by consumers that do not know them. The following quote from Hutter, Hautz, Dennhardt, and Füller (2013) shows the importance of trust before the use of a product: *“the consumer first attains awareness and knowledge about a product, subsequently develops positive or negative feelings towards the product and finally acts by buying and using or by rejecting and avoiding the product”*.

Because of the many decisions a buyer has to make before buying a product that is not known to them, consumers develop certain heuristics to help them get through the enormous amount of information offered (Jacoby, 1984; Scammon, 1977). Brands are one of these heuristics, offering the possibility to make a choice between a lot of products based on brand awareness rather than specific product details (Hutter et al., 2013). A brand is a cue of a product: it does not change the product itself, it only shows what a possible user might expect from the product before using it. The brand is accompanied by convictions about the quality of a product and can be a cue for trust.

Trust, being a crucial factor for product adoption, must be established by new products to enable mass adoption. Yamamoto and Lambert (1994) found evidence that product

appearance can exert an influence that can go beyond the influence of the performance or price attributes of that product when this information is not available. Other studies found that even in situations where there was sufficient information about the performance available and therefore influence of the design would not be expected, the same effect could be measured. (Madzharov & Block, 2010; Raghubir & Greenleaf, 2006; Townsend & Shu, 2010). It can be said that the trust in a product is also greatly influenced by the aesthetics of that product. Therefore, measuring trust based on aesthetics before the use of a product might be as important (or even more important) as measuring trust after the use.

2. Methods

2.1. Design of the survey

To design the digital version of the survey Qualtrics has been used, an online software package made by IBM to construct survey's. The design process of the survey in Qualtrics followed the original paper survey. Adaptation and design decisions have been explored through a mini-design review with Dr. Simone Borsci, to define the types of scales (Likert, Slider etc.) and to set up characteristics of the survey, such as: if-and cycles, sections and subsections, scenarios and randomization. Three scenarios were developed to be assigned to participants at random during the trials (Appendix C). The scenarios were designed to represent different real-life purposes for which a participant might want to buy a device. The survey is based on standardized questionnaires (McKnight et al., 2002) to measure general trust towards technology, general trust towards home MDDs and trust towards a specific device. Repeated measures have been used to enable the assessment of people's trust and decision making on the basis of a sequential set of information presented to the participants about four BPMs. Initially the participants had been offered only images about the devices. Gradually more information was added about the usability, aesthetics and other aspects (including previous user comments). The goal was to let participants reframe their decision to trust one of the four

BPMs.

Figure 1 presents the initial overall scheme of the paper survey (flow and functions). The final survey flow in Qualtrics is presented in Appendix D.

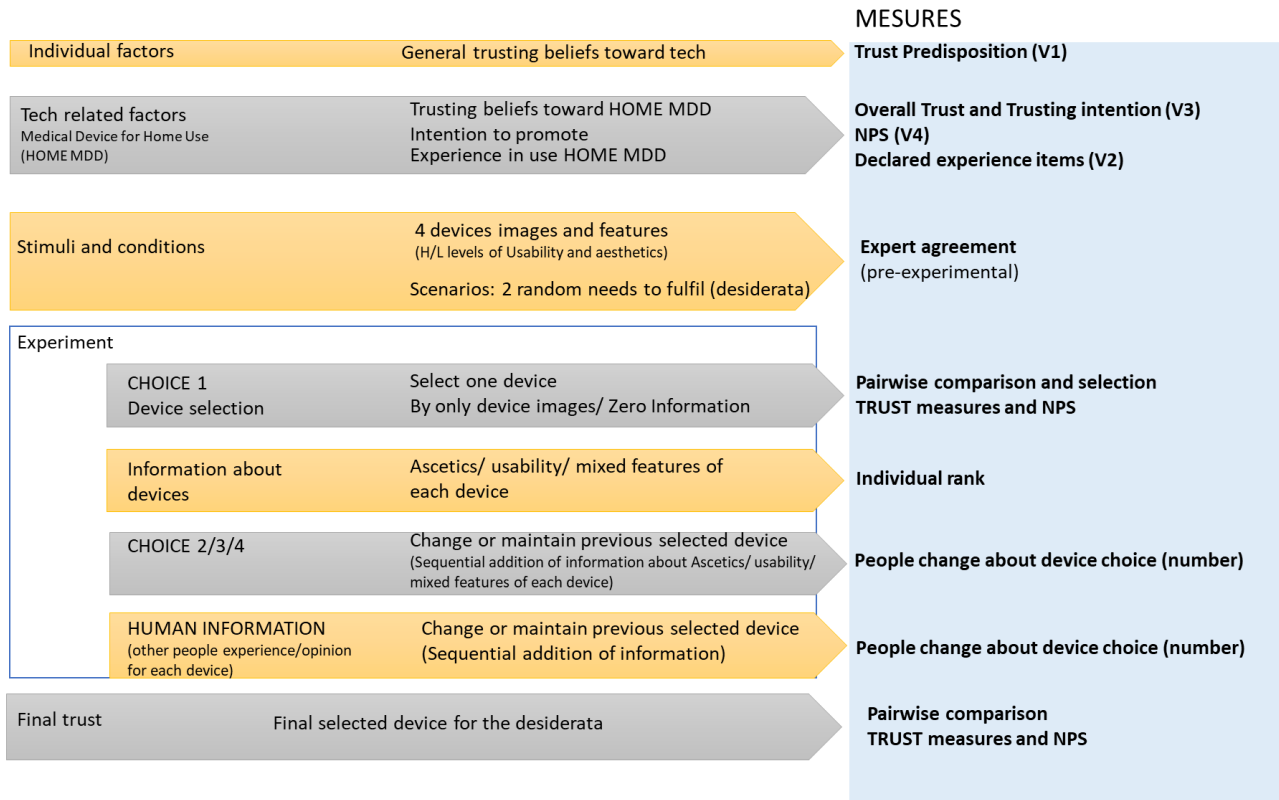


Figure 1. Paper survey flow and functions.

2.2. Expert agreement BPMs and definition of the stimuli

The initial selection of the four devices was performed to identify pictures and features of BPMs so that we may control the level of usability and aesthetics of the four devices. On the basis of our initial selections, we invited eight experts, throughout the mailing lists of the International Federation of Biomedical Engineers and UK NIHR London Diagnostic MIC. Experts' opinions were used to define the final list of characteristics for the four BPMs of the user survey, and to divide these characteristics in usability, aesthetics and mixed features. In that way we were able to control and manipulate the qualities (usability and aesthetics) of each device.

2.3. Usability evaluation of the digital survey

The usability of the survey was evaluated by using a retrospective thinking aloud method.

10 participants (Male: 4, Female: 6, Average Age: 23; SD: 2) have been recruited using the recruitment system of the University of Twente (Sona) and convenient sampling. Two participants applied for the study through Sona, the other eight participants have been recruited personally by the researcher. All participants were psychology students (bachelor and master) or psychology researchers.

The usability evaluation was performed in a closed lab environment at the University of Twente, and the BMS faculty. A Windows 10 computer was used to display the survey. A webcam and audio recording system was used to gather insights from the participants, and the screen actions were recorded using 'open broadcast software studio', a software package to record and broadcast. The recordings were stored on the hard drive of the computer for later analyses, and broadcasted to a private (invitation only) YouTube server, which ran on the researcher's student account of the Google user system provided by the university to each student and employee. For broadcast purposes the university's LAN network has been used. Each participant was given an informed consent form to sign (Appendix E). During the debriefing a NASA-TLX (Appendix F) was used to assess participants' cognitive workload during the interaction with the survey. The use of NASA-TLX in usability studies has a long tradition (Hart, 2006).

2.3.1. Procedure of usability evaluation

Each participant was asked to come to the lab at a given time, one at a time. The researcher would ask the participant to sit down in front of the computer and explain the purpose of the study. The researcher explained to the participant the main tasks: filling the survey and the NASA-TLX questionnaire, and inform the researcher about usability problems they encountered during a thinking aloud interview. The researcher also informed participants that their interaction and screen actions were video recorded, and monitored by the researcher computer,

to perform a post interaction retrospective thinking aloud (Janni Nielsen, Clemmensen, & Yssing, 2002; Van den Haak & De Jong, 2003).

During RTA participants were encouraged to provide any type of feedback (positive or negative) about their experience with the survey. The researcher asked as little as possible to not bias the participant. When the participant felt that everything was said, the researcher asked some questions based on notes made during the experiment.. When all points had been spoken of, the researcher would ask if there was something left the participant would like to add. After that the participant was thanked and the recording stopped. Recordings were numbered and saved according to the survey participant numbers for analysis.

3. Results

3.1. Expert analysis

The eight experts (3 Female; Average Age: 40,1) were asked to categorized each one of the 24 unique features into the following categories: usability, aesthetics, or a mix of both. Only six experts returned the categorization by mail. Table 1 gives an overview of the amount of features each expert put in each category. Table 2 gives a total overview of how often each feature has been categorized in each category.

Experts	Specification of BPMs		
	Usability	Mix	Aesthetics
1	8	10	6
2	13	8	3
3	3	21	0
4	10	13	1
5	6	12	6
6	3	15	6
Average	7,166667	13,16667	3,666667

Table 1. overall feature categorization for each expert

Category	BPMs Specifications	U	M	A	Selected
Usability	1. Device validation	3	3	0	
	2. Accuracy	4	2	0	X
	3. Device automatic detection of correct position of	3	3	0	

	arm/wrist for measurement				
	4. Cuff control	3	2	1	X
	5. Interface - Measurement initiation control	3	3	0	
	6. Extended results in the cloud	3	3	0	
	7. Irregularities report	3	3	0	
	8. Reminders about maintenance and calibration	3	3	0	
	9. Silent Mode	3	2	1	X
Mix	10. Interface - driven process of measurement	1	5	0	X
	11. Physical Memory	2	4	0	
	12. Smart connection/share information	2	4	0	
	13. Voice control interface	2	4	0	
	14. Display dimension and text visibility	0	6	0	X
	15. Unit of measure of BPM	2	4	0	
	16. Additional personalised results	0	6	0	X
	17. Battery types and charger	2	3	1	
	18. Communication modes of the device	0	6	0	X
Aesthetics	19. Information about cleaning	2	2	2	
	20. BPM weight and dimensions	0	2	4	X
	21. BPM display colour and definition	0	2	4	X
	22. Format of results	0	3	3	
	23. Case for device	2	1	3	X
	24. Case buttons and commands	0	3	3	

Table 2. The total amount each feature has been categorized in a category.

Each feature is added to the category in which it was mentioned most. When a feature was distributed evenly among multiple categories, it was added to the mix category. Eventually, the usability category contained nine features, the aesthetics category contained five features and the mix category contained 10 features. Of each category the features that had most mentions within that category were selected to compose our set of information about BPMs in the survey. As showed in Table 2, twelve features were selected out of the original list: 3 for usability, 4 mixed and 3 for aesthetics (see: Appendix G). This list of features was used to display to participants a description for each device, and to control for each one of the four BPMs a the Aesthetics and Usability levels as reported in Table 3. Participants were blind to the manipulation of the features and to the fact that the information were grouped in usability, aesthetics or mixed features.

BPMs	Manipulated features
1	Low aesthetics and usability

2	High usability and low aesthetics
3	High aesthetics and usability
4	High aesthetics and low usability

Table 3. Manipulation of usability and aesthetics per each BPM.

3.2. Usability Evaluation

3.2.1. Check of data export

The survey has been checked on the data it produces. Before the survey trial a simulation was ran. This is an internal function of the Qualtrics software package to check the survey flow and produces data. When exporting the data it became clear that the data file contained white spaces between lines of information. After unchecking the 'Remove line breaks' function while exporting the data this was fixed. The same process has been repeated after the trial of the survey. The data was correct after unchecking the 'Remove line breaks' option.

3.2.2. Usability issues

During the usability test a total of 29 issues have been mentioned by the participants and noted down by the interviewer (see Appendix H). The problems have been categorized by the researcher into categories that represent a general underlying problem in design. The categories have been based on Jakob Nielsen's heuristics for User Interface Design (Jakob Nielsen, 1995). Table 3 shows four categories in which we grouped 28 identified issues (two issues have been merged: 'visually hard and cluttered images' and 'cluttered text parts'). The categories have been ordered based on the severity for the survey flow or dataset. A link to the audio and video files and to the survey itself can be found in appendix I.

Category	Usability Issues
Flexibility and efficiency of use	1. Last question is reversed Likert scale
	2. List of statements: seem similar order, may lead to easy filling in and boredom
	3. Ranking order (unclear how to do it or not suitable)
	4. The list of statements is too long
	5. First time choosing the set of information: not totally clear/could be easier
Aesthetic and minimal design	6. First time showing scenario: not very visible
	7. Visually hard sometimes (text to close together, small text, small images, too many images)
	8. Background color (orange) can be annoying
	9. Last set of information seemed to be the same page as before
	10. The question numbering can be annoying

	11. Difference between choosing set of info and getting them: second one is easier to interpret
Consistency and standards	12. Spelling errors (i.e. upper harm, beliefs vs believes) and complicated sentences
	13. Scenario not very clear or relevant or well stated
	14. Degree level and associated years is not very clear
	15. Sets of information: features may be better explained
	16. Statements: it is said to use the scenario, but this is not always doable due to the stating of the statements
	17. Questions with long lines of answers
	18. Not all statements are easy to understand
	19. Thinking it is needed to give different answers at the statements due to new information given to the participants
	20. home MDD or wellbeing applications. Does this include phone apps?
	21. Sexes: also needs the option 'other'
	22. Not sure if everything should be in tune with the scenario
	23. Job: student worker, not clear
	24. Questions are not all exclusive, may be open for interpretation
	25. Not very professional: too much things that are not needed, sentences that are too long
Other related Issues	26. Configuration of the software: Qualtrics with auto translation may be a bad thing
	27. The distinction between the bpm's is not very clear
	28. First time images: participant feels like (s)he needs more information

Table 3. the categorization of the problems reported during the RTA interview.

3.2.3. Cognitive workload

The NASA-TLX showed a high mental demand and effort needed for completing the survey (Figure 2), however, as participant reported the instructions and the presentation of Qualtrics of the tool was unclear. As a consequences participants might have interpreted the scoring of scale in a wrong way. As participant (P001) explained: “[..] are the first words more to the left, and the last words more to the right? Or is it the higher the score, the more discouraged?”. He summarizes the issue as follows: “It is too open for interpretation”. Participant (P003) gives a clearer picture of the problem: “[You talk about] frustration and then you talk about insecure, discouraged, like all the things you associate with frustration; and then it’s about relaxed and stuff. But am I supposed to measure it from low relaxed to high relaxed or [..]”. Other two

participants reported similar issues in understanding how to assess the NASA-TLX. Because of this uncertainty within the NASA-TLX questionnaire on Qualtrics, we decided to report the findings but we are not going to consider this results to revise the survey usability.

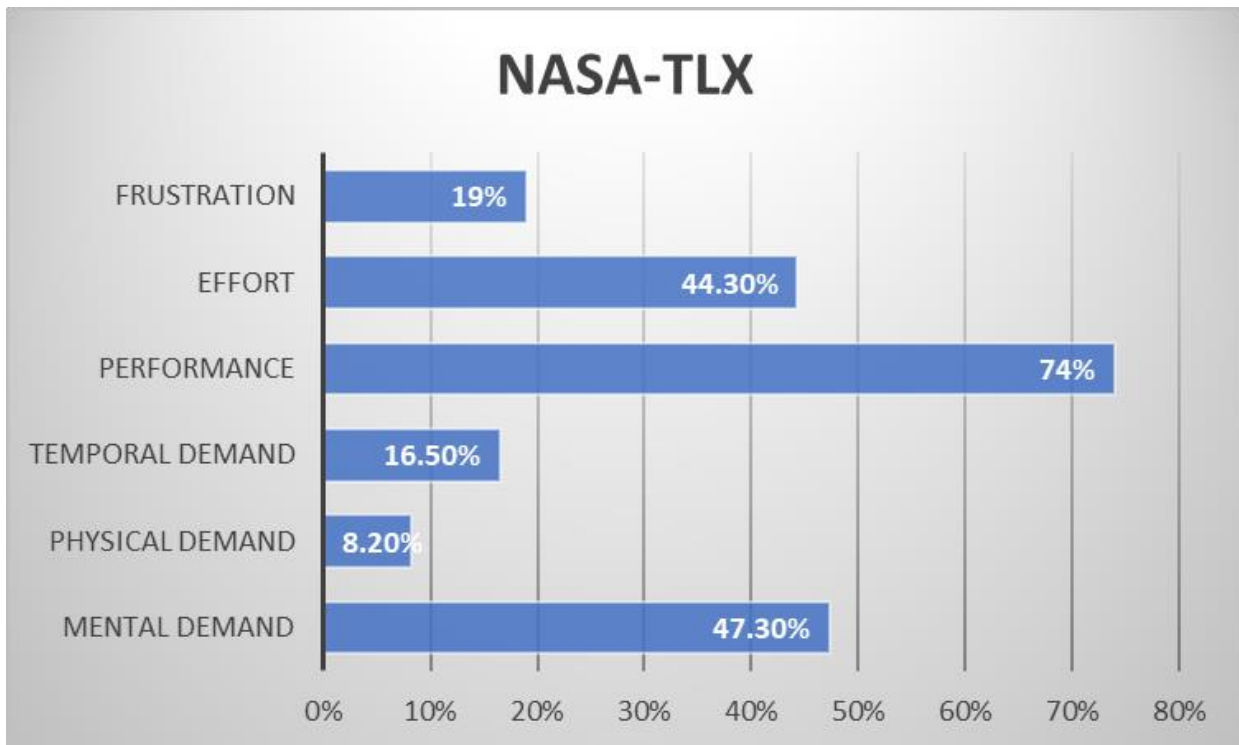


Figure 2. Overall scores of the NASA-TLX

4. Discussion

The current research succeeded in developing an usable survey to measure trust in the adoption of new technologies, specifically Home MDD's. Our test showed that data can be exported in (CSV format) in a clear way. The usability evaluation showed that despite each participant was able to complete the survey appropriately and without support, there are issues which need further inspection and redesign. We summarized below the key problems within each category and potential the recommendations for redesigning:

1. Flexibility and efficiency of use: This category contains problems that are applicable to a specific question that is not well designed, like a reversed Likert scale, a list of statements that is too long and that leads to uncaredful filling in of the list and not understanding the working of the ranking order. When a participant is not paying

attention to it, a reversed scales this may lead to biased results (*Recommendation 1*: revise and align the scoring of the questions in section code QEGT to all the other questions). Moreover, a list of statements is shown twice during the survey. Participants (2 out of 10) mentioned the list being too long, leading to a loss of focus. Others (4 out of 10) mentioned the list being the same both times, leading to carelessness in reading and answering the questions the second time (*Recommendation 2*: revise repeated statements QC1 and QFT1). Another important problem within this category is the ranking order of BPMs, which is mentioned by 4 out of 10 participants. The ranking order question seems to be hard to understand. Participants expected it would be possible to give different devices the same rank when they had the idea different devices where even in trustworthiness to them. The text explaining the working of the question is too small and contains too many unnecessary words (*Recommendation 3*: revised instruction of the ranking system for the device, in question QS1, and offer the possibility to participants to rank two devices are equal).

2. Aesthetic and minimal design: The overall appearance of the survey has been said to be professional, however, some parts need attention at the aesthetic design level. Most important are two pages that show new information, but appear to be the same as the page before, making participants miss the new information: the page when the scenario is shown first (page of QS1) and when the last set of information is added (section E). Participants mention that the images on the screen are the same which made them think nothing changed on the page. New information was added, but missed by the participants (*Recommendation 4*: Despite the need of a similar and coherent design, when new information is added to a page, it is necessary to draw the attention of the participants to the new data). Further problems within this category are mostly visual demanding parts of the survey. For example, some participants (5 out of 10) mentioned images being too cluttered or text being too small or close together. Another problem

that has been mentioned is the orange background when a question is selected. This is a build-in function of the Qualtrics software package (*Recommendation 5*: a general revision of the background-color and dimensions of the text may significantly reduce the participants' effort during the interaction).

3. Consistency and standards: The consistency and standards category can be divided into three different classes of problems: spelling and grammatical errors, missing options, and errors in stating questions. The first problem class is straightforward. The survey should be checked on spelling and grammar mistakes (*Recommendation 6*: Check the spelling and grammar of the survey). The second problem class, missing options, is applicable to multiple choice questions that seem to miss a certain option. Participants mentioned missing the option 'other' when selecting, for instance, their sex (D4) or when describing their job position (D5) (*Recommendation 7*: Ensure the option "other" is added where it is not currently implemented in the survey) . The last class, errors in stating questions, makes up a large part of the category. Participants mentioned that parts of the questionnaire are open for interpretation (i.e. the scenario is open for interpretation or unclear, statements are hard to understand, degree level and associated years are incorrect in the environment of the participant) or information is missing i.e. features that do not have a full description (*Recommendation 8*: arrange a second test, after redesign, asking to participants to evaluate, and maybe co-redesign unclear question statements.)
4. Other related issues: This category consists of the problems that have been mentioned by a single participant which are not very clearly stated (*Recommendation 9*: To extend the sample size and find out if these problems are persistent.)

4.1. Limitations and future work

The present work was only an initial attempt to build and pilot a digital instrument to evaluate

trust before the use of a medical device. To achieve our aim we faced two main challenges.

First, despite the precious work of the experts, another round of expert analysis is needed to reconsider the features list and to evaluate the qualities of each BPM. In fact, the section of the expert survey we used in Qualtrics to collect data about the professionals' opinion and trust toward each one of the BPMs was not well displayed, and output data were unusable. This was due to limitations of the in-built pair-wise comparison template of Qualtrics, which created confusion in the respondents. Fortunately, there was no problems in the part related to the categorization of the features. This experience helped us in the design of the end-user survey, however, this also left us without the part of the data we collected from the experts. Second, as we discussed in section 3.2.3, the digital version of NASA-TLX was perceived as problematic by participants. Future research needs to investigate how to appropriately collect data about the cognitive workload of participants of the survey. Finally, the main limitation of the present work, we would like to highlight that we involved in the usability test only a limited sample of students with a low level of expertise with medical device for home use, but with a good level of education. Future usability analysis, after the redesign, needs to include a larger and more differentiated sample of participants.

5. Conclusion

The final product (digital version of the survey) currently enables data collection and data exportation and analysis. The use of the digital tool will enable to reach people all over the world, especially because Qualtrics easily enables translation of the survey, to measure their trust before the use of devices for home use. Nevertheless, the present version of the digital survey needs a redesign and a retest before the tool could be considered a usable system. Nine actions that future researchers have to implement, are reported below, in terms of recommendations:

1. Revise the scoring of the questions in section code QEGT, and align to all the other questions in the survey;
2. Revise repeated statements QC1 and QFT1;

3. Revised instruction of the ranking system for the device, in question QS1, and offer the possibility to participants to rank two devices as equal;
4. Despite the need of a similar and coherent design, when new information is added to a page, it is necessary to draw the attention of the participants to the new data;
5. A general revision of the background-color and dimension of text may significantly reduce participant effort during the interaction;
6. A check of spelling and grammar of the survey;
7. Ensure the option “other” is added where it is not currently implemented in the survey;
8. Arrange a second test, after redesign, asking to participants to evaluate, and maybe co-redesign unclear question statements;
9. To extend the sample size and find out if these problems are persistent.

On the basis of the present analysis, the next phase of redesign and retest may aim to solve all the major issue in the survey and attempt to launch the survey for a large scale analysis after a second round of expert and usability evaluation.

Acknowledgment

International Federation of Biomedical engineers and UK NIHR London Diagnostic MIC for their help in recruiting experts. BMS Lab staff of Twente University for their support in performing the usability test.

References

- Aljazzaf, Z. M., Perry, M., & Capretz, M. A. (2010). *Online trust: Definition and principles*. Paper presented at the Computing in the Global Information Technology (ICCGI), 2010 Fifth International Multi-Conference on.
- Bond Jr, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and social psychology Review, 10*(3), 214-234.
- Bond Jr, C. F., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological bulletin, 134*(4), 477.
- Buchan, N. R., Croson, R. T., & Solnick, S. (2008). Trust and gender: An examination of behavior and beliefs in the Investment Game. *Journal of Economic Behavior & Organization, 68*(3-4), 466-476.
- Ekman, P., & Friesen, W. V. (1974). Detecting deception from the body or face. *Journal of personality and Social Psychology, 29*(3), 288.
- Fishwick, M. (2004). Emotional design: Why we love (or hate) everyday things. *The Journal of American Culture, 27*(2), 234-234.
- Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of Internet information credibility. *Journalism & Mass Communication Quarterly, 77*(3), 515-540.
- Flanagin, A. J., & Metzger, M. J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society, 9*(2), 319-342.
- Freud, S. (1905). *Introductory Lectures on Psychoanalysis*.
- Friedman, B., Khan Jr, P. H., & Howe, D. C. (2000). Trust online. *Communications of the ACM, 43*(12), 34-40.
- Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS quarterly, 27*(1), 51-90.
- Hansen, J. M., Saridakis, G., & Benson, V. (2018). Risk, trust, and the interaction of perceived ease of use and behavioral control in predicting consumers' use of social media for transactions. *Computers in Human Behavior, 80*, 197-206.
- Hart, S. G. (2006). *NASA-task load index (NASA-TLX); 20 years later*. Paper presented at the Proceedings of the human factors and ergonomics society annual meeting.
- Hutter, K., Hautz, J., Dennhardt, S., & Füller, J. (2013). The impact of user interactions in social media on brand awareness and purchase intention: the case of MINI on Facebook. *Journal of Product & Brand Management, 22*(5/6), 342-351.
- Jacoby, J. (1984). Perspectives on information overload. *Journal of consumer research, 10*(4), 432-435.
- Jarvenpaa, S. L., Tractinsky, N., & Vitale, M. (2000). Consumer trust in an Internet store. *Information technology and management, 1*(1-2), 45-71.
- Kelton, K., Fleischmann, K. R., & Wallace, W. A. (2008). Trust in digital information. *Journal of the Association for Information Science and Technology, 59*(3), 363-374.
- Kim, D. J. (2008). Self-perception-based versus transference-based trust determinants in computer-mediated transactions: A cross-cultural comparison study. *Journal of management information systems, 24*(4), 13-45.
- Lankton, N. K., & McKnight, D. H. (2011). What does it mean to trust Facebook?: examining technology and interpersonal trust beliefs. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems, 42*(2), 32-54.
- Lankton, N. K., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of the Association for Information Systems, 16*(10), 880.
- Lim, K. H., Sia, C. L., Lee, M. K., & Benbasat, I. (2006). Do I trust you online, and if so, will I buy? An empirical study of two trust-building strategies. *Journal of management information systems, 23*(2), 233-266.

- Lucassen, T. (2013). Trust in online information.
- Lynch, C. A. (2001). When documents deceive: Trust and provenance as new factors for information retrieval in a tangled web. *Journal of the Association for Information Science and Technology*, 52(1), 12.
- Madzharov, A. V., & Block, L. G. (2010). Effects of product unit image on consumption of snack foods. *Journal of Consumer Psychology*, 20(4), 398-409.
- McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information systems research*, 13(3), 334-359.
- Naef, M., & Schupp, J. (2009). Measuring trust: Experiments and surveys in contrast and combination.
- Nielsen, J. (1995, January 1, 1995). 10 Usability Heuristics for User Interface Design. Retrieved from <https://www.nngroup.com/articles/ten-usability-heuristics/>
- Nielsen, J., Clemmensen, T., & Yssing, C. (2002). *Getting access to what goes on in people's heads?: reflections on the think-aloud technique*. Paper presented at the Proceedings of the second Nordic conference on Human-computer interaction.
- Norman, D. A. (2004). Emotion design: Why we love (or hate) everyday things. In: Basic Books.
- Raghubir, P., & Greenleaf, E. A. (2006). Ratios in proportion: what should the shape of the package be? *Journal of Marketing*, 70(2), 95-107.
- Rocco, E. (1998). *Trust breaks down in electronic contexts but can be repaired by some initial face-to-face contact*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.
- Scammon, D. L. (1977). "Information load" and consumers. *Journal of consumer research*, 4(3), 148-155.
- Solomon, R. C. (2000). Trusting. *Heidegger, coping, and cognitive science: Essays in honor of Hubert L. Dreyfus*, 2, 229-244.
- Townsend, C., & Shu, S. B. (2010). When and how aesthetics influences financial decisions. *Journal of Consumer Psychology*, 20(4), 452-458.
- Van den Haak, M. J., & De Jong, M. D. (2003). *Exploring two methods of usability testing: concurrent versus retrospective think-aloud protocols*. Paper presented at the Professional Communication Conference, 2003. IPCC 2003. Proceedings. IEEE International.
- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*: John Wiley & Sons.
- Yamamoto, M., & Lambert, D. R. (1994). The impact of product aesthetics on the evaluation of industrial products. *Journal of Product Innovation Management*, 11(4), 309-324.

Appendix A. Initial paper concept of the trust survey

PHASE 2. SUREVEY DESIGN (TO BE REVISED WITH EXPERTS AND BY TRIAL)

[A group of people will look a multiple devices in the market (Hidden Brand, without a price limit) and they will be asked to look for information to suggest to another person to buy one product among alternatives.]

Surevey INTRO

SECTION DEMOGRAPHIC

D1. Age

D2. Nationality

D3. Gender

D5. Job position

D6. Degree

D6 We are interested in your general trust toward technologies, please rate your agreement to the question below

[[CONTROL VARIABLE 1] disposition to trust technology (7-point Likert scale from (1) strongly disagree to (7) strongly agree) (McKnight et al., 2002)]

1. My typical approach is to trust new technologies until they prove to me that I shouldn't trust them.
2. I usually trust in technology until it gives me a reason not to.
3. I generally give an technology the benefit of the doubt when I first use it.

[Page Break]

In particular, We are interested in the concept of trust toward medical devices for home use (HOME MDD):

This type of devices can be use by anybody to monitor themselves or by caregivers when a person can not use the device on their own.

HOME MDD include for instance: blood pressure monitor, pregnancy test, blood glucose monitors, sleep control device etc. [Inset pictures and explanation]

[CONTROL VARIABLE 2: EXPERIENCE inclusion/Exclusion criteria]

D7 I do have and use, or I used in the past a HOME MDD

Yes/No

D7.1. In the last 2 years, I have used one or more type of HOME MDD?

7-point Likert scale from (1) never (7) everyday

7.2. For instance, please indicate the device you had experience with:

[LIST]

[IF no to D7] D6.8 I have seen people in my family (or people I care of) use HOME MDD, and I have some knowledge about these tools

Yes/No

D8.1. In the last 2 years I have seen family members (or people I care of) use any type of HOME MDD?

7-point Likert scale from (1) never (7) everyday

D8.2 Please indicate the device you had an indirect experience with:

[LIST]

[IF NO to 7 and 8 EXIT participants Exclusion]

D9 Regarding HOME MDD in general please rate your agreement toward the statements below [TRUST intention (7-point Likert scale from (1) strongly disagree to (7) strongly agree) (McKnight et al., 2002)]

1. I feel people can depend HOME MDD.
2. People can always rely on results of HOME MDD when they need to use these tools.
3. I feel that people can count on HOME MDD when they need to use these tools.

D10 [NPS] (from 0 not at all likely to 10 extremely likely)

Think to the HOME MDD device you are more expert of, how likely you would recommend this device to a friend or a colleague of yours?

SECTION DEVICE PRESENTATION

In this survey we would like to discuss with you your trust about 4 different types of blood pressure monitors.

Blood pressure monitors are used ...[Describe use and purposes for different types of users]

In particular we present to you the following four devices:

DEVICE 1 (minimal set of information) , DEVICE 2 (minimal set of information), DEVICE 3 (minimal set of information), DEVICE 4 (minimal set of information)

e.g., Bestreviews.com/best-blood-pressure-monitors

SECTION DEVICE DESIDERATA

Each person 1 scenario [RANDOM]

(DESIDERATA 1) You need to buy a product for an elderly person who asked to you a favour to buy the best product for have accurate monitoring of his health condition at home.

(DESIDERATA 2) You need to buy a product for your family to serve the purposes of multiple people.

(CONTROL) You need to buy a product for you

SECTION A (CHOICE 1)

SA1. Imagine that you have no possibility to have other information about the device, just looking at the pictures of these devices how would do you rate the trust in technology (Intended as the believing that a technology has the desirable attributes for your needs) of each device:

Please compare the trust toward each device on the right toward each device on the left:

D1 Vs D2

D1 Vs D3

D1 Vs D4

D2 Vs D3

D2 Vs D4

D3 Vs D4

SA2. Overall, which is for you the most promising device i.e., the one who potentially seems more trustable:

D1/D2/D3/D4

SA3. Considering the device you selected as the most trustable among the options, please rate your agreement to the following questions:

[Technology trusting belief—functionality (7-point Likert scale from (1) strongly disagree to (7)

strongly agree) (McKnight et al., 2011)]

1. This device seems to have the functionality I need.
2. This device seems to have the features required for the desiderata.
3. This device seems to have the ability to do what I want it to do.

[Technology trusting belief—helpfulness (7-point Likert scale from (1) strongly disagree to (7)

strongly agree) (McKnight et al., 2011)]

4. I believe the end-user will not need help to use this device.

Technology trusting belief—reliability (7-point Likert scale from (1) strongly disagree to (7) strongly

agree) (McKnight et al., 2011)

6. This device seems very reliable.
7. I believe that this device will not fail me.
8. This device seems extremely dependable

[Trust Intention]

9. I feel people can depend on this device.
10. People can always rely on results of this device when they need to use it.
11. I feel that people can count on this device when they need to use these tools.

SA.4 [NPS] (from 0 not at all likely to 10 extremely likely)

Considering the device you selected, how likely you would recommend this device to a friend or a colleague of yours?

SA4.1. Why? (open questions)

SECTION B. CHOICE 2

You have now the possibility to have more information to better judge the products:

SB1. [Selection] Rank the usefulness of this information to inform your choice?

INFORMATION GROUP A: (Aesthetics features for experts)

INFORMATION GROUP B: (Aesthetics and usability features for expert)

INFORMATION GROUP C: (Usability and Functionalities features for experts)

CH1. BEFORE you have said that the more trustable device was [SA2]. Having this new set of information which one of the device seems more trustable in tune with the needs?

D1/D2/D3/D4

[IF selected device is Equal to SA2 go to SC1, IF different CH1.2]

CH1.2 Considering the device that seems more trustable please answer the question below:

Quest select from Lakantos 2015 technology, Humanness, and trust. See Reference and Adapt

SECTION C. CHOICE 3

SC1. [Selection] Which characteristics you want to know now (select one option)?

Remain 2 INFORMATION GROUP

[DIPLAY DEVICES WITH AND DISPLY NEW INFORMATION SELECTED]

CH2. BEFORE you have said that the more trustable device was [CH1]. Having this new set of information which one of the device seems more trustable in tune with the needs?

D1/D2/D3/D4

SECTION D. CHOICE 4

SD1. HERE YOU HAVE ADDITIONAL INFORMATION ABOUT THE PRODCUT FEATURES
[DIPLAY DEVICES WITH AND DISPLY NEW INFORMATION SELECTED]

CH3. BEFORE you have said that the more trustable device was [CH2]. Having this new set of information which one of the device seems more trustable in tune with the needs?
D1/D2/D3/D4

SECTION E. OTHER SAY INFO (CHOICE 5)

SE1. HERE YOU HAVE INFORMATION ABOUT WHAT OTHER PEOPLE SAID ABOUT THESE
DEVICES
[DIPLAY DEVICES WITH AND DISPLY NEW INFORMATION SELECTED]

CH4. BEFORE you have said that the more trustable device was [CH2]. Having this new set of information which one of the device seems more trustable in tune with the needs?
D1/D2/D3/D4

SECTION F. FINAL TRUST

You have selected the Device XX [for you/for the elderly person]

SF1. Please compare the trust toward each device on the right toward each device on the left:

D1 Vs D2

D1 Vs D3

D1 Vs D4

D2 Vs D3

D2 Vs D4

D3 Vs D4

SF1. Please considering the Device XX, rate your agreement toward the following statements
[Technology trusting belief—functionality (7-point Likert scale from (1) strongly disagree to
(7)

strongly agree) (McKnight et al., 2011)]

1.This device seems to have the functionality I need.

2. This device seems to have the features required for the desiderata.

3. This device seems to have the ability to do what I want it to do.

Technology trusting belief—helpfulness (7-point Likert scale from (1) strongly disagree to (7)
strongly agree) (McKnight et al., 2011)

4. The end-user will not need help to use this device.

Technology trusting belief—reliability (7-point Likert scale from (1) strongly disagree to (7)
strongly

agree) (McKnight et al., 2011)

5. This device seems very reliable.

6. I believe that this device will not fail me.

7. This device seems extremely dependable

[Trusting intention (7-point Likert scale from (1) strongly disagree to (7) strongly agree) (McKnight et al.,

2002)]

8. I feel people can depend on this device.

9. People can always rely on results of this device when they need to use it.

10. I feel that people can count on this device when they need to use these tools.

SF2 [NPS] (from 0 not at all likely to 10 extremely likely)

Considering the device you selected, how likely you would recommend this device to a friend or a colleague of yours?

SF2.1. Why? (open questions optional)

END SURVEY

Thanks

Appendix B. four BPMs and list of key features of BPMs



Feature 1 Device validation: Lay user are informed about the BPM validation and regulatory approval by device description, guidelines or packaging

Feature 2 Accuracy: Lay users are informed about the validated levels of accuracy of the BPM results by the device description, guidelines or packaging.

Feature 3 Device automatic detection of correct position of arm/wrist for measurement: Lay users are informed if the BPM may or may not automatically control the body position of the end-users, and advice the user on how to position their body to correctly execute the measurement.

Feature 4 Cuff control: Lay user are informed if the device may or may not detect the cuff positioning and advice about the correct positioning and tightness of cuff.

Feature 5 Interface - driven process of measurement: Lay users are informed if the BPM may or may not drive with (textual/graphics) explanation the end-user to perform correctly the sequential steps for the measurement.

Feature 6 Interface - Measurement initiation control: Lay users are informed if the BPM may or may not suggest to the end-users to relax for 5 minutes before the performance of (a non continuous) measurement to ensure reliable one off measurements.

Feature 7 Physical Memory: Lay users are informed about the available memory of the BPM in terms of number of recorded readings that may be stored by the device

Feature 8 Extended results in the cloud: Lay users are informed if the BPM offers or not the possibility to store and visualised data in a cloud by using a mobile app.

Feature 9 Smart connection/share information: Lay users are informed if the BPM has or not a WiFi connection to share or print results when needed.

Feature 10 Voice control interface: Lay users are informed if the BPM **offers or not** a voice control option to enable end-users to perform certain tasks without dealing with the digital interface

Feature 11 BPM weight and dimensions: Lay users have information (associated to pictures) about the weight and dimension of the BPM

Feature 12 BPM display colour and definition: Lay users are informed about the definition of the display and if the results are reported with colour or black and white digits/graphical elements.

Feature 13 Display dimension and text visibility: Lay users are informed about the dimension of the digital display, and about the visibility of the digits i.e., accessibility and readability

Feature 14 Format of results: Lay users are informed if the BPM reports results only by numbers or by also including graphical presentations.

Feature 15 Unit of measure of BPM: Lay users are informed if the device report measure is mmHg or kPa, and if the BPM offers or not the possibility to the end users to select the type of unit they want to visualise.

Feature 16 Irregularities report: Lay users are informed if the BPM may or may not report irregularities of people heartbeat, or irregularities due to technical faults

Feature 17 Additional personalised results: Lay users are informed if the BPM **offers or not** the possibility to visualise person based results, such as end-users weekly average against normal value

Feature 18 Case for device: Lay users are informed if a case to safely transport the BPM is included or notwith the device

Feature 19 Battery types and charger: Lay users are informed about the type of battery of the BPM and if an adapter/charger is or is not included with the device

Feature 20 Communication modes of the device: Lay users are informed if the BPM provides only visual-text messages or also other signals (auditory, lights and graphical) to support end-users interaction and to communicate errors/problems

Feature 21 Reminders about maintenance and calibration: Lay users are informed if BPM may or not advice to end-user when and how to perform software updates and re-calibration tasks.

Feature 22 Information about cleaning: Lay users are informed if the BPM may or not support with an interface-driven process task related to cleaning and disinfecting the device

Feature 23 Case buttons and commands: Lay users are informed if the BPM have commands on the case which are visible, appropriately text label and if buttons have a back-light to enhance visibility

Feature 24 Silent Mode: Lay users are informed if the BPM offers or not a silent option to minimise the noise and perform a quiet measurement in any environment.

Appendix C. Scenario's used during the test

Scenario 1

PLEASE SELECT A BPM IN TUNE WITH THIS SCENARIO:

<< Tom, who is 65 years old, asked you to do a favour to help him identify the best product to support him with accurate monitoring of his health condition at home>>

Scenario 2

PLEASE SELECT A BPM IN TUNE WITH THIS SCENARIO:

<<Some your family members would like to buy a BPM for different reasons, to control their well-being but also for sport training purposes. They asked you to identify a device who may fit their different needs>>

Scenario 3

PLEASE SELECT A BPM IN TUNE WITH THIS SCENARIO:

<<You would like to buy a BPM for yourself, and you would like to find a device which has good features to monitor your well-being>>

Appendix D. Survey Flow in Qualtrics

Standard: Consent Form (3 Questions)

Branch: New Branch

If

If By clicking "I agree" below you are indicating that you are at least 18 years old, have read and... No, I do not want to consent Is Selected

EndSurvey: Advanced

Standard: Demographic (7 Questions)

Standard: Inclusion/Exclusion (8 Questions)

Standard: Home MDD General (3 Questions)

BlockRandomizer: 3 - Evenly Present Elements

EmbeddedData

Condition = 0

EmbeddedData

Condition = 1

EmbeddedData

Condition = 2

Branch: New Branch

If

If Condition Is Equal to 0 bytes

Standard: Condition 0 (Elderly) (1 Question)

Branch: New Branch

If

If Condition Is Equal to 1 bytes

Standard: Condition 1 (Family) (1 Question)

Branch: New Branch

If

If Condition Is Equal to 2 bytes

Standard: Condition 2 (family) (1 Question)

Block: Scenario and Rank (5 Questions)

EmbeddedData

Rank = \${q://QID79/ChoiceGroup/ChoiceWithLowestValue}

Standard: Post Choice question (4 Questions)

Standard: F CHOICE 0 (3 Questions)

Standard: F CHOICE 1 (5 Questions)

EmbeddedData

Q22_answer = \${q://QID91/ChoiceGroup/SelectedChoices}

Standard: F CHOICE 2 (12 Questions)

EmbeddedData

Q25_answer = \${q://QID104/ChoiceGroup/SelectedChoices}

Standard: F CHOICE 3 (5 Questions)

EmbeddedData

Q27_answer = \${q://QID111/ChoiceGroup/SelectedChoices}

Standard: Including reviews (2 Questions)

EmbeddedData

Q29_answer = \${q://QID122/ChoiceGroup/SelectedChoices}

Standard: Questions and choice (7 Questions)

Standard: Block 15 (2 Questions)

Appendix E. The informed consent

NOTE:

This survey was optimized for computer screen or large screen tablets. Although it could be accessed by mobile phone, some of the questions and the elements of the survey could be less accessible and usable than expected with small screens.

Information Sheet and consent form

Introduction

You are being invited to participate in a preliminary study done by Bjorn Wesselink (b.h.a.wesselink@student.utwente.nl) supervised by Dr Simone Borsci (s.borsci@utwente.nl) from the University of Twente .

We are investigating the concept of trust toward diagnostic medical device for home use (HOME MDD).

In particular we are interest in people **trust on technology before the use**, which could be described as **a set of believes and expectations people have about the ability of a new technology (its functionalities and characteristics) to fit appropriately their needs.**

Purpose of the study

The purpose of this preliminary phase is to gather your opinion and trust before use about four types Blood Pressure Monitors (BPM) for home use and to improve the survey as a tool for a large scale research. In this phase we are looking for the usability of the survey and the comprehensibility of the questions and text.

You will be asked to fill the survey, and you will be interviewed about it. Our **aim is to gather data to understand your trust expectations toward new technology and medical device for home use, but also to improve the quality of interaction with the survey.**

Sections of the survey, time to complete and procedure

The survey will take you approximately between **10 to 15 minutes** to be completed.

We will record your screen activity and facial expressions, and we will audio record your comments.

After the completion of the survey:

- 1- I will ask you to rate the effort in filling the online survey, with a NASA-TLX questionnaire.
- 2- we will go through the screen recording together to analyze issues or problems you experienced with the survey I will ask you to verbalize any problem you faced related to the text and image comprehensibility, or its interactive functions. Remember that we are interested in your opinion. Both negative and positive comments will help us to redesign and improve the survey.

Rights of participants

There are no right or wrong answers in this study, we are interested in your opinion and you have the right to express positive and negative comments about the survey, or ask to the interviewer any questions you feel are important for you to better understand the tasks.

Moreover, you have the right to quit the experiment at any time.

Risks and data management

We believe there are no known risks associated with this study; however, as with any online related activity the risk of a breach of confidentiality is always possible. To the best of our ability your answers in this study will remain confidential and anonymized, and data will be secured and stored in an encrypted repository.

Use of data

Anonymized data will be used for statistical and research purposes, and data analysis will be used for reports and scientific publications.

Contacts

Your participation in this study is completely voluntary and you can withdraw at any time.

If you have any questions concerning your rights as a participant to this study, **you may contact Dr Simone Borsci (s.borsci@utwente.nl)**

Please print a copy of this page for your records. Thank you!

Appendix F. NASA-TLX

Q1 Participant ID (for the interviewer)

Q2 We would like to have insights about the mental workload you spent in filling the survey. Please for each of the 6 scales below rate, form 0 (LOW) to 100 (HIGH), how demanding was the survey

	0	10	20	30	40	50	60	70	80	90	100
1. MENTAL DEMAND (How much mental and perceptual activity was required to fill the survey e.g., thinking, deciding, remembering, looking, searching, etc.? Was the task easy or demanding, simple or complex, exacting or forgiving?) ()											
2. PHYSICAL DEMAND (How much physical activity was required e.g., pushing, pulling, turning, controlling, activating, etc.? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?) ()											
3. TEMPORAL DEMAND (How much time pressure did you feel to fill the survey? Was the task slow and leisurely or rapid and frantic?) ()											
4. PERFORMANCE (How successful do you think you were in filling the survey? How satisfied were you with your performance in accomplishing the goal?) ()											
5. EFFORT (How hard did you have to work mentally and physically to fill the survey appropriately?) ()											
6. FRUSTRATION (How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed did you feel during the completion of the survey?) ()											

Appendix G. Final features of the devices displayed in the survey

Accuracy	Acceptable	High accuracy	High accuracy	Acceptable
Cuff control Device detects whether the cuff is wrapped tightly enough	No	Yes	Yes	No
Silent mode Device offers a silent option to ensure quiet measurement in any context	No	Yes	Yes	No
Driven measure Device suggests body, arm and cuff position to ensure reliable measures	No	Yes	Yes	No
Display visibility text digits size and visibility of results	Quite acceptable visibility of results and digits	Quite acceptable visibility of results and digits	High visibility of results and large digits	High visibility of results and large digits
Personalised results Automatic display of data personalised based on your weekly measures.	No	Yes	Yes	No
Auditory interface Auditory signals are given when the device starts and finalise the measure, or when errors occurred.	No, only text	No, only text	Yes	Yes
Weight & dimensions	Acceptable weight and portability	Acceptable weight and portability	Light and compact (extremely portable)	Light and compact (extremely portable)
Colour screen & definition	No colour and low defintion	No colour and low defintion	Colour with high definition	Colour with high definition
Case for device A case is provided when you buy the device	No	No	Yes	Yes

Appendix H. Usability issues

Reported errors	Times reported
1. Spelling errors (i.e. upper harm, beliefs vs believes) and complicated sentences	9
2. Last question is reversed Likert scale	8
3. First time choosing the set of information: not totally clear/could be easier	6
4. Scenario not very clear or relevant or well stated	5
5. Ranking order (unclear how to do it or not suitable)	4
6. Degree level and years is not very clear	4
7. List of statements: seem similar order, may lead to easy filling in and boredom	4
8. Sets of information: features may be better explained	4
9. Background color (orange) can be annoying	3
10. First time showing scenario: not very visible	3
11. The picture thing: too many and too small. Also not clear for the current subject. Images in general	3
12. Questions with long lines of answers	2
13. The list of statements is too long	2
14. Statements: even when requested to use the scenario, some say what do you think of.. Is unclear	2
15. Visually hard sometimes (text to close together, small text)	2
16. Q9.3 home MDD or wellbeing applications. Does this include phone apps?	1
17. Sexes: also the option 'other'	1
18. Job: student worker, not clear	1
19. First time images: participant feels like (s)he needs more information	1
20. The question numbering can be annoying	1
21. Last set of information seemed to be the same page as before	1
22. Questions are not all exclusive, may be open for interpretation	1
23. The distinction between the bpm's is not very clear	1
24. Not all statements are easy to understand	1
25. Thinking it is needed to give different answers at the statements due to new information given to the participants	1
26. Not very professional: too much things that are not needed, sentences that are too long	1
27. Configuration of the software: Qualtrics with auto translation may be a bad thing	1
28. Not sure if everything should be in tune with the scenario	1
29. Difference between choosing set of info and getting them: second one is easier to interpret	1

Appendix I. The informed consent

Link to the audio and video files:

https://drive.google.com/drive/folders/1y_2aL42RFKpEMaM8Jc3VNas0r5PwsVg?usp=sharing

Link to the designed survey:

https://utwentebseu.qualtrics.com/jfe/form/SV_dcmB2JxNiyIRNcx

Link to the NASA-TLX:

https://utwentebseu.qualtrics.com/jfe/form/SV_dhBXX6VZW4TDbIp