Inclusive mHealth design: Usability of the Dutch Talking Touch Screen Questionnaire: Qualitative study

Welbie M. ¹, Wittink H. ¹, Westerman M.J.², Topper I.¹, Snoei J.¹, Devillé W.L.J.M.³⁴⁵

1. Research Group Lifestyle and Health, Research Center Healthy and Sustainable Living, Utrecht University of Applied Sciences, Utrecht
2. VU University, Institute of Health Sciences, Amsterdam Public Health research institute, Department of Methodology and Statistics, Amsterdam, the Netherlands
3. Julius Centre for health Sciences and Primary Care, University Medical Centre Utrecht, The Netherlands
4. Faculty of Social and Behavioral Sciences, University of Amsterdam, Amsterdam, the Netherlands
5. National Knowledge and Advisory Centre on Migrants, Refugees and Health (Pharos), Utrecht, the Netherlands

Journal: JMIR mHealth uHealth

Number of text pages of the entire manuscript (including figures and tables): 33

Number of tables: 7

Number of figures: 0 = all figures can be found in Multimedia Appendix 1

Word count 5832

The author to whom correspondence should be sent:
Marlies Welbie MSc. P.T.
Hogeschool Utrecht
Lectoraat Leefstijl en Gezondheid
T.a.v. Marlies Welbie MSc. P.T.
Postbus 12011
3501 AA Utrecht
Telephone number: +316 38192100
E-mail address: marlies.welbie@hu.nl
http://www.onderzoek.hu.nl/Onderzoekers/Marlies-Welbie.aspx
ABSTRACT

Background
In the past years a mHealth app, called the Dutch Talking Touch Screen Questionnaire (DTTSQ) was developed in The Netherlands. The aim of development was to enable Dutch physical therapy patients to autonomously complete a health related questionnaire regardless of their level of education or digital literacy.

Objective
The aim of this study was to evaluate the usability, defined as the effectiveness, efficiency and satisfaction, of the prototype of the DTTSQ for Dutch physical therapy patients with diverse levels of education and experience in using mobile technology.

Methods
The qualitative Three-Step Test-Interview method, including both think-aloud and retrospective probing techniques, was carried out to get insight into the usability of the DTTSQ. Twenty-four physical therapy patients were included. The interview data were analyzed using a thematic content analysis approach aimed at analyzing the accuracy and completeness with which participants completed the questionnaire (effectiveness), the time it took the participants to complete the questionnaire (efficiency) and the extent to which the participants were satisfied with the ease of use of the questionnaire (satisfaction). The errors encountered by the participants in this study were given a severity rating which was used to provide a rough estimate of the need for additional usability efforts.

Results
All participants within this study were very satisfied with the ease of use of the DTTSQ. Nine participants stated that the usability of the application exceeded their expectations. The group of four average/high experienced participants encountered only one error in total, while the eleven little experienced participants encountered an average of 2 errors per person and the nine non-experienced participants an average of 3 errors per person. Thirteen different kind of errors were found during this study. Three of these errors need to be addressed before the DTTSQ will be released because they have the potential to negatively influence future usage of the tool. Ten errors were less likely to influence future usage of the tool substantially.

Conclusions
The usability of the DTTSQ needs to be improved before it can be released. No problems were found with satisfaction or efficiency during the usability-test. The effectivity needs to be improved by 1. increasing the visibility of activated answers, 2. making it easier to navigate through screens and 3. not forcing respondents to answer questions that do not apply to their situation. This study shows the importance of including less skilled participants in a usability study when striving for inclusive design. Further research is necessary to gain more insight into the needs, preferences, capacities, values, and goals in relation to mHealth technology of people with low literacy skills, low educational levels and no or little experience with using mobile technology.

Key words: mHealth; eHealth; surveys and questionnaires; physical therapy specialty; qualitative research
INTRODUCTION

EHealth is developing rapidly [1]. It is defined as the use of information and communication technology (ICT) in healthcare [2]. A growing amount of literature indicates that using eHealth can improve the accessibility, quality and efficiency of healthcare [3,4,5]. It seems to be effective for people who have access to it and are able to use it well, which is not the case for everybody [6,7]. For instance, people with low income or low education and people who are 65 years and older are vulnerable when it comes to effective eHealth use. In these populations access to the internet and hardware like personal computers, tablets, mobile phones and smartphones and experience and skills to use these devices is low [6,7,8,9]. Differences between people regarding digital skills and access to internet and hardware is often referred to as the digital divide [10,11]. Since eHealth technologies are usually primarily developed for people who are experienced and skilled in using ICT [12,13], people who do not have access to ICT or are not skilled in using it, are at risk of being excluded from the use of eHealth. Looking at the widespread expansion of eHealth technologies this encompasses the potential threat of contributing to the ongoing exacerbation of health inequalities in western countries [1]. However, if the needs, preferences, capacities, values, and goals of potential future users who do not have good access to internet and digital technology or who are not well skilled in using this technology, would be explored and taken into account during each stage of development of eHealth tools, eHealth could potentially reduce health inequalities [14].

The development of a specific form of eHealth technology, called mobile health (mHealth) technology, seems especially promising when it comes to reducing health inequalities [15,5,16,17]. MHealth has been defined by the Global Observatory for eHealth of the World Health Organization as “medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices” [18]. A recent project called eSalud showed that mHealth can be cost effective, help to overcome cultural and language barriers and provide health information and services to low health access areas [15]. Furthermore, recent publications indicate that the digital divide is narrowing because of the increased ownership of mobile devices such as smartphones and tablets [5,16,17].

Still, having access to internet and digital technology does not automatically mean that people are able and willing to use it effectively to increase their health nor that different people use it in the same way [14, 19, 20,21,22,23,24,25]. Recent studies found ethnic and socioeconomic differences in mHealth usage [19,20] and it is known that older people use mHealth differently from younger people [14]. And though the gap of people owning tablets and smartphones between groups is closing, still a substantial amount of people do not own such devices. For instance the percentage of Dutch citizens of 65 years and older owning a tablet computer in 2017 was 55.2% versus 75.8% citizens of 12-25 years of age [26]. Considering that vulnerable groups, like people with low income and low education, bear a disproportionate burden of disease [27,28] and the amount of healthcare visits increases with age [29] it is to be expected that a relatively large amount of care recipients do not have a lot of experience using mobile technology. To fulfill the promise of mHealth technology contributing to reduction of health inequalities, it is very important to carefully test the usability of mHealth applications in research populations which include members of the target populations that are at risk of being excluded from usage of the tested tool.
In the past years a prototype of a mHealth application, called the Dutch Talking Touch Screen Questionnaire (DTTSQ), was developed in The Netherlands. The aim of this development was to enable Dutch physical therapy patients to autonomously complete a user-friendly health-related questionnaire regardless of their level of education or digital literacy. Because it is not to be expected that all physical therapy patients own a tablet computer, the DTTSQ is meant to be presented in a physical therapy practice on a tablet computer that is owned by the physical therapy practice concerned. Patients are asked to complete the DTTSQ in the waiting room of the physical therapist prior to their first visit. The development of the prototype of the DTTSQ, which runs on a tablet computer, was described in detail by Cremers et al. in 2015 [30]. Prior to the current study the prototype was only tested in a sample with low education outside of the physical therapy context.

The aim of this study was to test the prototype of the DTTSQ within the physical therapy context to see what parts of the prototype needed adjustment in order for it to be user-friendly for physical therapy patients regardless of their level of education and level of experience with operating mobile technology.

The research question underlying this study was: What is the usability of the prototype or the DTTSQ for physical therapy patients with different levels of education and experience in using mobile technology?
METHODS

Design
A qualitative descriptive study was carried out. Observational data on the way participant’s operated the DTTSQ was collected through The Three-Step Test-Interview (TSTI) method [31]. This method includes both think-aloud and retrospective probing techniques.

Definitions
**Usability** was defined by the International Standards Organization (ISO) as “the effectiveness, efficiency and satisfaction with which specified users can achieve goals in particular environments” [32]. **Effectiveness** is the accuracy and completeness with which users achieve certain goals [33]. In this study error rates and severity of errors were used as the primary indicator of effectiveness. **Efficiency** is the relation between the accuracy and completeness with which users achieve certain goals and the resources expended in achieving them [33]. In this study task completion time was used as an indicator of efficiency. **Satisfaction** is the users’ comfort with and positive attitudes towards the use of a system [33]. In this study participants were interviewed about their satisfaction with the ease of use of the DTTSQ. Ease of use was defined as the degree to which the usage of a particular system is free from effort [34].

Setting and Participant Selection
Data was collected in the same study population and at the same time as the data reported in an paper earlier published by Welbie et al. [35]: Recruitment took place in eleven primary care practices in deprived areas of Utrecht, The Netherlands. Potential participants were invited by their physical therapists to participate in this study. The physical therapists shortly explained the goal of the study and provided the patient with an information letter that was written in plain Dutch language. If patients were interested, the physical therapist asked permission to give the patients’ telephone number to researcher IT. Then researcher IT (1) Contacted the patient by telephone, (2) Again shortly explained the aim of the study, (3) Made sure the patient understood what was asked of him/her, (4) Answered any question the potential participant may have had, and (5) Checked the inclusion criteria. Inclusion criteria for participants were as follows: aged 18 years or older, Dutch as their first language, and both parents born in The Netherlands. The sampling procedure was aimed at getting a broad variation in levels of education and age, plus balance in our sample regarding gender. Throughout the recruitment process, the recruiting physical therapists were constantly kept informed about the profiles of participants the researchers were looking for. In total, 24 physical therapy patients were included in this study [35]. Characteristics of study population can be found in tables 1 and 2.

<table>
<thead>
<tr>
<th>Table 1. Characteristics of study population</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Characteristics</strong></td>
</tr>
<tr>
<td>Mean age (range) years</td>
</tr>
<tr>
<td>Gender, n (% of N)</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
</tbody>
</table>
Level of education, n (% of N)

<table>
<thead>
<tr>
<th>Level of Education</th>
<th>N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>6 (25)</td>
</tr>
<tr>
<td>Moderate</td>
<td>13 (54)</td>
</tr>
<tr>
<td>High</td>
<td>5 (21)</td>
</tr>
</tbody>
</table>

Self-declared experience with using mobile technology, n (% of N)

<table>
<thead>
<tr>
<th>Experience with Mobile Technology</th>
<th>N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>9 (37)</td>
</tr>
<tr>
<td>Little</td>
<td>11 (46)</td>
</tr>
<tr>
<td>Average/high</td>
<td>4 (17)</td>
</tr>
</tbody>
</table>

*Low = no education or primary education
Moderate = lower secondary education, (upper) secondary education or post-secondary non-tertiary education (including vocational education)
High = Tertiary education (bachelor’s degree or higher)

Table 2. Characteristics per respondent

<table>
<thead>
<tr>
<th>Pseudonym</th>
<th>Experience Mobile Technology</th>
<th>Age</th>
<th>Level of Education</th>
<th>Last Occupation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ida</td>
<td>None</td>
<td>66</td>
<td>low</td>
<td>cleaning lady</td>
</tr>
<tr>
<td>Bill</td>
<td>None</td>
<td>72</td>
<td>moderate</td>
<td>order picker</td>
</tr>
<tr>
<td>Mia</td>
<td>None</td>
<td>73</td>
<td>moderate</td>
<td>administrative officer</td>
</tr>
<tr>
<td>Dora</td>
<td>None</td>
<td>77</td>
<td>low</td>
<td>cleaning lady</td>
</tr>
<tr>
<td>Ilene</td>
<td>None</td>
<td>79</td>
<td>low</td>
<td>cleaning lady</td>
</tr>
<tr>
<td>Bob</td>
<td>None</td>
<td>68</td>
<td>moderate</td>
<td>cashier</td>
</tr>
<tr>
<td>Jerome</td>
<td>None</td>
<td>47</td>
<td>low</td>
<td>truck driver</td>
</tr>
<tr>
<td>Helga</td>
<td>None</td>
<td>54</td>
<td>high</td>
<td>artist</td>
</tr>
<tr>
<td>Michelle</td>
<td>None</td>
<td>56</td>
<td>low</td>
<td>cleaning lady</td>
</tr>
<tr>
<td>Roger</td>
<td>Little</td>
<td>70</td>
<td>moderate</td>
<td>home painter</td>
</tr>
<tr>
<td>Peter</td>
<td>Little</td>
<td>18</td>
<td>moderate</td>
<td>student</td>
</tr>
<tr>
<td>Christine</td>
<td>Little</td>
<td>39</td>
<td>moderate</td>
<td>nurse of mentally disabled people</td>
</tr>
<tr>
<td>Jill</td>
<td>Little</td>
<td>55</td>
<td>high</td>
<td>management assistant</td>
</tr>
<tr>
<td>Lydia</td>
<td>Little</td>
<td>56</td>
<td>moderate</td>
<td>domiciliary care</td>
</tr>
<tr>
<td>Rose</td>
<td>Little</td>
<td>60</td>
<td>moderate</td>
<td>sales woman</td>
</tr>
<tr>
<td>Francine</td>
<td>Little</td>
<td>61</td>
<td>moderate</td>
<td>administrative officer</td>
</tr>
<tr>
<td>Harald</td>
<td>Little</td>
<td>63</td>
<td>high</td>
<td>financial controller</td>
</tr>
<tr>
<td>Henry</td>
<td>Little</td>
<td>64</td>
<td>moderate</td>
<td>project coordinator</td>
</tr>
<tr>
<td>Ronald</td>
<td>Little</td>
<td>70</td>
<td>low</td>
<td>home painter</td>
</tr>
<tr>
<td>Bernie</td>
<td>Little</td>
<td>76</td>
<td>high</td>
<td>lecturer chemistry</td>
</tr>
<tr>
<td>Jude</td>
<td>average/high</td>
<td>18</td>
<td>moderate</td>
<td>student</td>
</tr>
<tr>
<td>Joline</td>
<td>average/high</td>
<td>19</td>
<td>moderate</td>
<td>photographer</td>
</tr>
<tr>
<td>Ellen</td>
<td>average/high</td>
<td>32</td>
<td>high</td>
<td>management assistant</td>
</tr>
<tr>
<td>Sandra</td>
<td>average/high</td>
<td>39</td>
<td>moderate</td>
<td>graphic designer</td>
</tr>
</tbody>
</table>

*Low = no education or primary education
Moderate = lower secondary education, (upper) secondary education or post-secondary non-tertiary education (including vocational education)
High = Tertiary education (bachelor’s degree or higher)
Content of the Dutch Talking Touch Screen Questionnaire
The prototype of the Dutch Talking Touch Screen Questionnaire was a digital application on a tablet computer. It was developed during a co-design process [36] which in this case meant that a group of ten low educated people helped to design the questionnaire. As a result of the co-design process, questions on pain location and pain intensity were added to the original questions of an existing questionnaire which aims to select limitations in functioning and to formulate specific treatment goal(s) [37,38]. Furthermore, visual (videos and pictures) and auditory (speech technology) support were added to enable participants to see and hear the questions which were shown on separate screens. Response items could be selected by tapping on the touch screen and plain language was used in all spoken and written text within the Dutch Talking Touch Screen Questionnaire [30]. An overview of all types of screens is given in the Multimedia Appendix 1. The eight questions of the questionnaire can be found in Multimedia Appendix screenshots 2, 3, 4, 7, 9, 11, 12 and 13.

Instructions
Instructions were given in the form of three video clips:
1. an introduction clip in which the purpose of the questionnaire and all functions of the questionnaire were explained (see Multimedia Appendix screenshot 1).
2. an instruction clip in which the purpose of question 4 and a newly added navigation function were explained (Multimedia Appendix screenshot 6).
3. a closing clip in which the respondent is thanked, explained what the physical therapist would do next and told that the questionnaire would close down automatically (Multimedia Appendix screenshot 16).

Functions
Next button:
Navigation function to go to the next screen. Not activated unless a response item is selected (except for question 4 (see Multimedia Appendix screenshot 7)).
Help button:
Activates the help function: the text on the screen is read aloud, the purpose of the question is explained and operating instructions for the particular screen are provided.
Correction function:
Tapping a second time on a response item de-selects the item.
Stop button:
Escape function: shuts down the questionnaire. All previous given answers are saved.

Overviews
To help respondents keep track of their answers, overviews of previous given answers were provided regularly during completion of the questionnaire (see Multimedia Appendix screenshots 5, 8, 10, 14 and 15).
Multimedia Appendix 1

Screenshot 1 ‘Welcome’
Introduction movie

Screenshot 2 ‘Pain’
Question 1: “Do you have pain? Yes/No”
**Screenshot 3 ‘Location of the health problem’**
Question 2: “Tap on the location of your health problem. You can tap on multiple locations.”

**Screenshot 4 ‘pain severity’**
Question 3: “This is the location of your pain. Rate the severity of your pain on the scale below.”
Screenshot 5 ‘Overview location of the health problems’
Overview answers question 1-3:
“This is the location of your health problems.”

Screenshot 6 ‘Activities’
Instruction movie question 4

Liggen
Druk op de actie ‘Liggen’ als klachten heeft.
Screenshot 7 Activity ‘lying’
Question 4: “Select the activities in which you are impaired”

Activiteiten

Liggen
Druk op de activiteiten waarbij u klachten heeft.

On this screen you see all the activities that you selected in previous screens. These are the activities in which you are impaired.

Screenshot 8 ‘Overview activities’
Overview answers question 4:
“On this screen you see all the activities that you selected in previous screens. These are the activities in which you are impaired.”

Activiteiten

U ziet op het scherm alle activiteiten die u in de vorige schermen gekozen heeft.
Dit zijn alle activiteiten waarbij u klachten heeft. Klopt dit?
**Screen 9 ‘Most important activities’**

**Question 5:** “Select the three activities which are most important to you”

Belangrijkste activiteiten

Kies de 3 activiteiten die u het meest belangrijk vindt.

- Reiken
- Iets voortduwen
- Aankleden of uitkleden
- Haren kammen
- Oprapen

**Screen 10 ‘overview most important activities’**

Overview answers question 5: “You chose these three activities. Is this correct?”

Belangrijkste activiteiten

U koos deze 3 activiteiten. Klopt dit?

- Raiken
- Aankleden of uitkleden
- Haren kammen
**Screenshot 11 'Most important activity 1'**
Question 6: “Select the activity which is most important to you”

Belangrijkste activiteit 1

Kies de activiteit die het belangrijkst is voor u.

- Reiken
- Aankleden of uittrekken
- Haren kammen

**Screenshot 12 'Most important activity 2'**
Question 7: “Which of these two activities is still most important for you now?”

Belangrijkste activiteit 2

Welke van deze twee activiteiten is nu voor u nog het belangrijkste?

- Reiken
- Haren kammen
Screenshot 13 ‘Effort activity 1’

Question 8: “Rate the effort it takes to carry out this activity”

Moeite activiteit 1

Geef op de balk aan hoeveel moeite deze activiteit u kost.

0 1 2 3 4 5 6 7 8 9 10

STOP Verder

Screenshot 14 ‘overview most important activities and effort’

Overview answers question 6-8: “On this screen you see the activities that are most important to you in order of most important to least important. Is this correct?”

U ziet nu op het scherm de activiteiten die voor u het belangrijkst zijn op volgorde van meest belangrijk naar minst belangrijk. Klopt dit?

6 8 9

Aankleden of uitkleden Haren kammen Reiken

STOP Verder
Overview answers total questionnaire:
“On the screen you see an overview of all your answers you provided until now.”

Screenshot 15 ‘overview all outcomes of the questionnaire’
Overview answers total questionnaire:
“On the screen you see an overview of all your answers you provided until now.”

Screenshot 16 ‘Thank you’
Closing movie
Data collection and procedures
Data collection took place at the physical therapy practice or the participant’s home, depending on the preference of the participant. Researchers IT and JS were present. Researcher IT was in the lead during the interviews. Researcher JS asked complementary questions if she missed information.

The following steps were taken according to the TSTI method [31]:

Step 1
Each participant was observed by researchers IT and JS while they were completing the DTTSQ thinking out loud. This step was aimed at collecting observational data regarding the usability of the DTTSQ. The data collected consists of two kinds: 1. observations of respondent’s behavior and 2. think-aloud data. The data was recorded on videotapes as well as audiotapes. Additionally, the researchers took real time notes for use during the following steps of the interviews as well as for later analysis. The researchers wrote their notes down on hardcopies of print screens of the Dutch Talking Touch Screen Questionnaire. Researchers IT and JS noted problems with operating the tablet computer including using the touchscreen, navigating through the questionnaire, understanding the task given in each screen, selecting response items and using the correction function. They also wrote down when the stop button was used.

Step 2
Researcher IT conducted an in-depth interview after the respondent finished completing the DTTSQ. Data collection during this step was exclusively focused on filling possible gaps and checking the observational data collected during step 1.

Step 3
During step 3 of the TSTI researcher IT conducted a semi-structured interview aimed at eliciting experiences and opinions of the participant. Researcher JS was allowed to ask complementary questions, if she felt it was necessary, in order to get complete and rich data. When participants had encountered problems in operating the DTTSQ they were asked what they thought the exact nature and possible cause of each type of problem was and how they tried to overcome the problem. Then the participants were questioned about their satisfaction regarding the ease of use of the user interface, technical operation, layout and content and overall usability of the DTTSQ. The participants were stimulated to report feelings, express opinions, preferences and recommendations. Researcher IT finished the interview by collecting demographic data and data on self-reported experience with mobile technology (see table 1 and 2).

Analyses
Data were analyzed using a thematic content analysis approach [39]. Four types of data were analyzed: 1. Video recordings of the completion of the questionnaire, 2. field notes of the observed respondent behavior 3. transcriptions of the audio recordings of the semi-structured interviews and 4. background information regarding the educational level, age, gender, and self-reported experience with using mobile technology.
To get more familiar with the data and to create an overview, researcher MW made a
descriptive summary of each case on the basis of all four types of generated data. Each
summary contained information on whether or not the questionnaire was fully completed,
if, when and why the stop function was used, the kind of problems that occurred with the
operation, the completion time and all emerging themes regarding satisfaction or
dissatisfaction with the ease of use of the questionnaire. The summaries were
supplemented with information regarding educational level, age, gender and experience in
using mobile technology.

Subsequently, researcher MW derived the observed problems from the summaries. She
clustered the errors. Every new problem was categorized as a new error. MW analyzed the
video recordings to see how many times each error was made in total, per participant and
per question/screen of the questionnaire. After a full overview of errors had emerged she
categorized the errors into one of the four levels of severity described by Nielsen and
Loranger [40]: low, medium, serious or critical. To score severity she used the method of
Hattink et al. [41]: The severity was scored by answering the three questions of Nielsen and
Loranger [40] with 'yes' (= one point) or 'no' (= 0 points):
1. Frequency: Does a substantial amount of users encounter the problem?
2. Impact: Does the problem cause much trouble to those users who encounter it?
3. Persistence: Does the problem cause trouble repeatedly?
This resulted in a 0-3 point score per error. Each score was related to a level of severity: 0 =
low, 1 = medium, 2 = serious and 3 = critical.
These severity-ratings give an indication of which errors lead to disastrous usability problems
and which errors are more cosmetic in nature [42]. This provides insight into whether or not
the usability of the DTTSQ needs to be improved before it can be released. Nielsen and
Loranger recommend to tackle only serious and critical severe errors during the
development process of a digital tool. Low and medium severe errors do not have priority
according to Nielsen and Loranger, because although they are bothersome, they are not
likely to directly influence the usage of a tool. This makes it uninteresting to tackle them
from a cost-benefit perspective. Serious and critical severe errors on the other hand can be
so disrupting that they can make users stop using a tool or prevent them from even starting
to use it at all. Therefore they should not be ignored during the development process of a
digital tool [40].

As a next step researcher MW started open coding of all fragments in the transcripts of the
semi-structured interviews that were related to (dis)satisfaction about the ease of use of the
questionnaire using Maxqda 10 of VERBI Software GmbH, Berlin. After she finished open
coding she organized and structured the codes until a coding scheme emerged on the basis
of which the part of the research question that was related to satisfaction of the participants
could be answered sufficiently.

As a last step researcher MW ordered the analyzed data into three groups: data of
participants who had 1.no, 2. little and 3. average/high experience in using mobile
technology. This was done to see whether or not data differed within and between these
groups.
During the whole course of the study procedures and results were checked and discussed with researchers HW, MJW and WD.

**Ethics**

No external funding was received by the Utrecht University of Applied Sciences to conduct this study. This study was submitted to the Medical Ethics Commity of the Academic Medical Centre of Amsterdam which declared that it does not fall under the scope of the ‘Medical Research Involving Human Subjects Act’. The study was conducted according to the principles of the Declaration of Helsinki. All participants provided written informed consent. The participants’ names used in this article are all fictitious in order to protect their privacy.
RESULTS

Effectiveness
Nine out of the twenty-four participants in this study did not complete the DTTSQ fully (see table 3). Michelle, Bill and Helga stopped completing the questionnaire by using the stop button. Ida, Ilene, Dora, Peter, Rose, and Mia went through the whole questionnaire but unintentionally left one or more parts open.

Table 3. Experience with mobile technology and completion of the DTTSQ

<table>
<thead>
<tr>
<th>(sub)population</th>
<th>not fully completed</th>
<th>fully completed</th>
</tr>
</thead>
<tbody>
<tr>
<td>No experience using mobile technology</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Little experience using mobile technology</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Average/high experience using mobile technology</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td>Total population (N=24)</td>
<td>9</td>
<td>15</td>
</tr>
</tbody>
</table>

Unanswered (parts of) questions
Michelle, Ida, Ilene, Dora, Peter, Rose and Mia failed to fully complete the DTTSQ because they failed to select answering-options and/or unintentionally skipped questions by double-tapping on the next button (see error 1-5 in table 4). All participants, except for Michelle, additionally failed to notice they had not effectively selected an answer because the difference between activated and non-activated answers was not accentuated enough (see error 6 in table 4).

Use of the stop button
When Michelle noticed most of her answers were missing from the summary in question 6, she got confused. In question 6 she was asked to choose the three most important activities in which she was limited. The screen contained only one activity-photo while she, in her mind, had selected a lot of photo’s earlier. Except for the one photo that she had managed to select, she had tapped on the text beneath the photos, in which case the item was not activated (see error 5 in table 4). The activity on the one photo that she had managed to select was of no priority to her. Therefore she decided to use the stop button and ended the questionnaire.

Bill had a lot of trouble operating the questionnaire. He commented on the introduction clip: “I do not think that what she is saying is difficult, but I just am not able to remember it. I have no experience with these kind of devices. So I forgot what she said right away.”

Bill managed to get to question 4 by activating the help function on each screen he entered. When he touched the navigation button to see all the activity-photo’s in question 4, the photo gallery moved in a different direction then he had presumed. This startled him somewhat and made him forget that he had to push the next button to go to the next screen (see error 7 in table 4). He activated the help function again, but that was of no use anymore. After trying a few buttons without succeeding to go to the next screen he gave up and tapped on the stop button.
Helga operated the digital questionnaire fluently until she had to choose the three activities that were most important to her in question 5. She did not use the navigation function of the photo-gallery and as a result she did not see all her earlier selected activities (see error 4 in table 4). She chose the three most important activities out of the five photos that were immediately visible. When she realized what happened she wanted to pause for a moment to find out how she could change her answer. She interpreted the stop button as a ‘time-out-function’ and was a bit shocked when she found out that she had stopped the questionnaire altogether.

A complete overview of frequency and severity of all errors encountered can be found in table 4.

<table>
<thead>
<tr>
<th>Error/problem</th>
<th>Amount of Participants</th>
<th>Frequency</th>
<th>Severity rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Accidentally skipping a screen by double tapping on the next button</td>
<td>8</td>
<td>16</td>
<td>medium</td>
</tr>
<tr>
<td>2. Double-tap on answering option causing activation and deactivation of the answer of choice</td>
<td>1</td>
<td>1</td>
<td>Low</td>
</tr>
<tr>
<td>3. Skipping a screen by accidently touching the next button with the palm of the hand</td>
<td>1</td>
<td>1</td>
<td>Low</td>
</tr>
<tr>
<td>4. Not using the navigation function of the photo-gallery in question 4 causing the respondent not seeing all presented response items</td>
<td>2</td>
<td>2</td>
<td>Low</td>
</tr>
<tr>
<td>5. Touching the text underneath a photo in question 4 to select an activity instead of touching the photo itself causing the activity not to be selected</td>
<td>3</td>
<td>30</td>
<td>Low</td>
</tr>
<tr>
<td>6. Not able to see whether or not a selected answer is activated (not accentuated enough)</td>
<td>8</td>
<td>8</td>
<td>serious</td>
</tr>
<tr>
<td>7. Not knowing how to get to the next screen</td>
<td>1</td>
<td>1</td>
<td>serious</td>
</tr>
<tr>
<td>8. Pushing too hard or tapping too soft on the touch-screen causing the touch screen not to respond</td>
<td>11</td>
<td>40</td>
<td>medium</td>
</tr>
<tr>
<td>9. Not able to correct a wrong answer</td>
<td>8</td>
<td>13</td>
<td>Medium</td>
</tr>
<tr>
<td>10. Not reading the text above the photos of question 5 causing the respondent to keep on performing the task given with question 4</td>
<td>4</td>
<td>8</td>
<td>Low</td>
</tr>
<tr>
<td>11. Not noticing that the multiple NRS-effort scores in question 8 are related to different activities, which by mistake results in identical scores for different activities</td>
<td>1</td>
<td>1</td>
<td>Low</td>
</tr>
<tr>
<td>12. Scoring the bodychart in question 2 mirrored</td>
<td>2</td>
<td>2</td>
<td>Low</td>
</tr>
<tr>
<td>13. Scoring (serial) questions that do not apply to the respondents' situation (forced by the software)</td>
<td>1</td>
<td>4</td>
<td>Serious</td>
</tr>
</tbody>
</table>
Severity of errors
The scores that led to the severity rating can be found in table 5. Error six, seven and thirteen scored ‘serious’ and all other errors scored medium or low which means that they do not need to be addressed in further development of the DTTSQ [32]. No error scored ‘critical’.

Table 5. Severity rating for errors according to Nielsen and Loranger [40] using the scoring method of Hattink et al. [41]

<table>
<thead>
<tr>
<th>Error/problem</th>
<th>Frequency score</th>
<th>Impact Score</th>
<th>Persistence Score</th>
<th>Total Score</th>
<th>Severity Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Medium</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Low</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Low</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Low</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Low</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>Serious</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>Serious</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Medium</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Medium</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Low</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Low</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Low</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>Serious</td>
</tr>
</tbody>
</table>
Amount of errors

Ellen, Jill, Lydia, Christine, Sandra and Joline were able to complete the questionnaire without any errors. The other eighteen participants were not able to operate the questionnaire fluently. Participants with no experience in using mobile technology encountered an average of 3 errors per person, participants with little experience encountered 2 errors on average per person and only one error occurred within the study sample that had average/high experience using mobile technology (see table 6).

Table 6. Amount of errors per level of experience with using mobile technology

<table>
<thead>
<tr>
<th>Error</th>
<th>No experience n=9</th>
<th>Little experience n=11</th>
<th>Average/high experience n=4</th>
<th>Total population N=24</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>3</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>4</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>5</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
</tr>
</tbody>
</table>
**Efficiency**

The twenty-one participants who got to the end of the questionnaire had an average completion time of ten minutes and twenty-five seconds. Non-experienced participants needed more time than little experienced participants did, who in their turn needed more time than average/high experienced participants did (see table 7).

<table>
<thead>
<tr>
<th>(sub)Population</th>
<th>Mean completion time</th>
<th>Median completion time</th>
<th>Range of completion times</th>
</tr>
</thead>
<tbody>
<tr>
<td>No experience with mobile technology (n=6)</td>
<td>11.38 min</td>
<td>9.38 min</td>
<td>8.2 min – 22.10 min</td>
</tr>
<tr>
<td>Little experience with mobile technology (n=11)</td>
<td>10.41 min</td>
<td>9.57 min</td>
<td>6.54 – 18.10 min</td>
</tr>
<tr>
<td>Average/high experience with mobile technology (n=4)</td>
<td>7.55 min</td>
<td>7.42 min</td>
<td>5.50 min – 10.26 min</td>
</tr>
<tr>
<td>Total population (n=21)</td>
<td>10.25 min</td>
<td>9.43 min</td>
<td>5.50 min – 22.10 min</td>
</tr>
</tbody>
</table>

**Satisfaction**

All participants were satisfied with the ease of use of the questionnaire. The use of plain language, the way ICT was used and the way the user interface was designed was greatly appreciated by the participants:

Dora:

> "Everything was well described. I am not always able to understand everything, but this went well. I understood what was asked of me."

Roger:

> "I have trouble operating my mobile phone and I own a notebook but don’t you ask me how that thing works! I am capable of a lot but I am not technical in that way. [...] This was the first time for me to use a tablet computer. I only had to follow the instructions. I did not have to start it up or open something, it just started working and it shut down by itself. I thought it was easy to work with. Better than when you have to write things down."

Ellen:

> "I am a very visual person. And this thing is very visual. [...] Like green is ‘no pain’ and red is ‘a lot of pain’.”

All participants were satisfied with the completion time of the Dutch Talking Touch Screen Questionnaire.

**Satisfied despite encountered problems**

Operation problems, regardless of the amount and severity of the errors encountered by each individual participant, did not influence satisfaction about the ease of use of the questionnaire. Francine, for instance, was asked how she felt about the fact that the application did not always respond to her touch right away (see error 8 in table 4). She encountered this problem thirteen times in total. She lightheartedly answered:
"Oh these are things that happen. I experience the same things with my own computer. My computer refuses to sometimes, so... I think I was just pushing too hard on the tablet sometimes, that’s all.”

When Bill, who used the stop button, was asked if he would have preferred a paper-based questionnaire he said:

“No. It took me some time to get used to it but it is easy to use actually.”

Expectations exceeded
Nine participants explicitly stated that operating the questionnaire was easier than they had expected beforehand. When Ida was confronted with the questionnaire she agitatedly said:

“Never in a million years I believe I can do this. That I can tell you right away.”

Noticeably reluctant and nervous she started to complete the questionnaire. When she finished she seemed surprised and relieved. She smiled and said:

“Okay? So this was the questionnaire? [...] Ooooh but this was doable! I thought I would have to look up things and operate it like my grandchildren do.”

And then she started laughing out loud and cheerfully asked if anyone would like to have some coffee.

Christine was positively surprised too:

“It responds really well. Normally I am not that good with screens, but this is easy. It almost feels like a game! It really responds nicely. Nothing disappears when I touch it. It reacts very calmly but at the same time it is very fast. I really like that it contains photo’s instead of drawings. It is instantly clear: these are my activities and that is what they mean by ‘sitting down’. You see it right away. I also like the regular summaries. It keeps you on track and enables you to check whether or not you forgot something.”

Participants’ recommendations for improvement
The most mentioned recommendations for improvement of the usability of the DTTSQ by participants were: shorten the length of the instructions, accentuate the activated response items and improve the user interface of question 4 by giving respondents a complete overview of activities to choose from in one screen, without having to use complicated navigation functions.
DISCUSSION

Principal Results

All participants within this study were very satisfied with the ease of use of the DTTSQ. Nine participants stated that the usability of the application exceeded their expectations. The participants who had no experience with using mobile technology completed the prototype of the DTTSQ less effectively and efficiently than the little- and average/high experienced participants did. In the group of average/high experienced participants only one error was encountered in total, while the non-experienced participants encountered an average of three and the little experienced an average of two errors per person. Thirteen different kinds of errors were encountered during this study. From a cost-benefit perspective three of these errors will need to be addressed during future development of the DTTSQ, because they have the potential to influence the future usage of the tool negatively [40]. The three errors that need to be addressed are: error 6 ‘not able to see whether or not a selected answer is activated (not accentuated enough)’, error 7 ‘not knowing how to get to the next screen’ and error 13 ‘scoring (serial) questions that do not apply to the respondents’ situation (forced by the software)’. Participants also recommended to shorten the length of the instructions and improve the user interface of question 4 by giving respondents a complete overview of activities to choose from in one screen, without having to use complicated navigation functions.

Comparison with Prior Work

Frokjaer et al. consider effectiveness, efficiency and satisfaction as independent aspects of usability [33]. The results of the current study confirm that it is risky to assume that there are correlations between these aspects [33]. All participants, including participants who were not able to fully complete the questionnaire because of problems they had with operating the application, were satisfied with the usability of the DTTSQ. Looking solely at the high satisfaction with the ease of use of all participants one could make the assumption that the DTTSQ is, usability-wise, ready to be released. Looking at the data found on efficiency within this study one can see that more experienced participants need less time to complete the questionnaire. This seems logical and the completion-time was acceptable to all participants. Based on the efficiency results solely one could therefore also conclude that the DTTSQ was ready to be released. Looking at the effectiveness-results and specifically at the severity-rates of the errors that occurred during the response-process, the researchers of the current study concluded that the usability of the DTTSQ needs to be improved to prevent error 6, 7 and 13 from occurring before it can be released. Looking at the results of all three tested aspects of usability within the current study the researchers concluded that although the DTTSQ needs some improvement to prevent three serious errors from occurring, the potential of the DTTSQ as a user-friendly tool for all physical therapy patients regardless of their experience with using mobile technology and their educational level, is promising.

According to Frokjaer et al. relations between the three aspects of usability depend in complex ways on the application domain, use context and user’s experience [33]. User’s experience may well have been of influence on the satisfaction outcomes of the current study. Eighty-three percent of the total study population had no or little experience in using
mobile technology (see table 1 and 2). Limited or no user experience may have caused a
form of computer anxiety, resulting in low self-efficacy, which in its turn led to low
expectations towards the ease of use of the DTTSQ [43]. Nine out of the twenty-four
participants in the current study explicitly stated that operating the DTTSQ was easier than
they had expected beforehand. The other participants did not explicitly state this, but their
statements on the ease of use could easily be interpreted as such. No participant stated or
gave the impression that the ease of use of the DTTSQ was lower than they would have
expected. According to the Expectation Confirmation Theory [44] actual performance
exceeding the expectations of testers leads to satisfaction among these testers. The more
their expectations are exceeded the more satisfied testers will become. Due to the limited
user experience of most of the study participants, expectations towards the ease of use of
the DTTSQ may have been low, which may have made it easier to exceed them. Especially
considering that the DTTSQ was specifically designed to be easy to use for low educated
people who lack the necessary skills to use ICT [30].

Comparable studies have reported often that study participants were very satisfied with the
ease of use of a tested tool, while data on efficiency and effectivity of the use of that tool
showed that there was room for improvement [45–48]. The patient population
characteristics of three of these studies are comparable with the current study [45,46,48].
Two of these studies had a qualitative component in their study design. The qualitative
results of both these studies also show that participants’ expectations regarding the ease of
use of the tested tool were exceeded [46,47]. It is reasonable to assume that limited user
experience may have led to low expectations regarding the ease of use of the tested tools in
these studies and therefore played a role in their high satisfaction outcomes.

Strengths and limitations
It is a strength of this study that non as well as little, average and high experienced users of
mobile technology were included. Although recommended in the literature [12,49], to this
date there has been an insufficient amount of empirical studies to prove the worth of
involving future users at risk of exclusion in the development process of eHealth tools [50].
In a recent review Latulippe et al. found only three studies that involved future users at risk
of exclusion in their design and evaluation processes [8]. This paper contributes to the body
of knowledge of inclusive mHealth design which involves active participation of vulnerable
potential future users in usability evaluation.

Data of the current usability study were collected in parallel with the data on the response
processes which were published earlier by Welbie et al [35]. Insight into the response
processes of the participants gave a first impression of the face validity of the DTTSQ.
Because usability issues can influence the response process, it is a strength of this study that
both forms of data were collected at the same time within the same research population.
The results of both studies should be taken into account simultaneously during the further
development of the DTTSQ. The usability and the response processes will have to be re-
tested in exactly the same manner after adjustments in the DTTSQ have been made. This
process will have to be repeated until an acceptable level of usability and face validity of the
DTTSQ is reached.
For the data collection of both the current as the response process study the qualitative TSTI method [31] was chosen. The current study shows that this research method is suitable for collecting data on the response processes as well as on usability. In addition this method suited the needs of low educated participants by not demanding any reading or writing skills from them. A downside of the chosen method is the lack of generalizability of the data. After the DTTSQ will reach an acceptable level of usability and face validity, the next step in research should be quantitative usability-, validity- and reliability testing producing generalizable data. Selecting quantitative methods for this purpose may become quite challenging in striving for inclusive design. Researchers will be challenged to develop a quantitative study design which will enable people with low literacy skills or low educational levels to participate. Research designs that include reading and writing tasks for participants may lead to exclusion of these vulnerable and hard to reach populations [51].

Conclusions
The usability of the DTTSQ needs to be improved before it can be released. No problems were found with satisfaction or efficiency during the usability-test. Effectivity needs to be enhanced by 1. increasing the visibility of activated answers, 2. making it easier to navigate through the screens and 3. not forcing respondents to answer questions that do not apply to their situation. Participants additionally recommended to minimize the length of the instructions and present all the answering options of question 4 in one screen.

Considering the difference in amount of errors encountered by non and little experienced participants versus average/high experienced participants within this study, it can be concluded that in striving for inclusive design it is vital to involve potential future users at risk of exclusion in usability testing.

The results of this study also emphasize the importance to always make sure that effectiveness and effectivity and satisfaction are investigated during usability studies. This may be even more important in usability studies that include participants that are at risk of exclusion of future usage of the tested tool. Researchers who want to investigate the usability of mHealth tools in a population that includes participants that have no or little experience in using mobile technology should take into account that these participants may lack prior experience or even have negative experiences with using mHealth tools which may result in low expectations of the usability of such tools. Satisfaction outcomes are influenced by the expectations participants have towards the ease of use of the tested tool before the test. It could be interesting to measure and further investigate computer-anxiety and self-efficacy towards the use of the tested tool prior to and after usability testing in order to be able to put satisfaction outcomes into perspective. Furthermore, positive satisfaction outcomes of usability tests should always be presented in combination with outcomes of efficiency and effectivity in order to give the recipient a complete and realistic overview of the usability of the tool.

Further research is necessary to gain more insight into the needs, preferences, capacities, values, and goals in relation to mHealth technology of people with low literacy skills, low educational levels and no or little experience with using mobile technology. Insight is also needed into what effects meeting these user requirements will have on the future use of these tools by these specific populations.
Acknowledgments
The authors would like to thank Anita Cremers PhD for her contribution to reflecting on the data from a designer perspective and Kees van der Veer PhD and Jan Pool PhD for helping them to think about the application of the TSTI method in low-educated populations.

Conflicts of Interest
None declared.
REFERENCES


