Think-aloud usability testing of an app to increase physical activity

Tillmann Fischer (✉ tillmann.fischer@tu-dresden.de)  
Technische Universitat Dresden Medizinische Fakultat Carl Gustav Carus  
https://orcid.org/0000-0003-3411-9838

Gesine Reinhardt  
Technische Universitat Dresden Medizinische Fakultat Carl Gustav Carus

Paul Stumpf  
Technische Universitat Dresden Medizinische Fakultat Carl Gustav Carus

Peter E.H. Schwarz  
Technische Universitat Dresden Medizinische Fakultat Carl Gustav Carus

Lorenz Harst  
Technische Universitat Dresden Medizinische Fakultat Carl Gustav Carus

Research article

Keywords: App, usability, think-aloud, smartphone, physical activity, test

Posted Date: December 13th, 2019

DOI: https://doi.org/10.21203/rs.2.17034/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background Insufficient physical activity is one of the most important risk factors for non-communicable diseases and should therefore be intensively prevented in all age groups. According to several trials, activity can be effectively increased by smartphone-based interventions. However, the use of digital applications depends on many factors such as user acceptance and intuitive operability. Therefore, usability testing has proven to be important for the successful development of digital interventions. Thus, the German app “VIDEA bewegt” was tested for its usability in order to assess strengths and weaknesses and to improve the app. Methods In April 2019, ten interviews were conducted using the think-aloud method and following a standardised protocol. The adult participants were confronted with “VIDEA bewegt” for the first time and had to perform pre-defined tasks covering the main features of the app. Demographics of the participants and their expectations towards the app were collected. The interviews were recorded and transcribed. The analysis of the transcripts was performed independently by two team members following the deductive qualitative content analysis according to Mayring. Results The app was rated positively in terms of design, basic registration process and the largely self-explanatory navigation. Users missed an explanation of the basic structure and included components of the programme. Several usability problems were described, including technical errors and problems with understanding individual control elements. Four of ten participants could imagine using the app in the future. Conclusions All test participants were able to use the app independently to a large extent and most of the a priori defined goals were achieved successfully. With certain minor changes, the usability can therefore be assessed as good. Trial registration The performed usability test is the basis of a study named “Evaluation of an app-based activity intervention for statutory health insured people” which is registered in the German Clinical Trials Register (DRKS). DRKS-ID: DRKS00017392 (14 June 2019)

Background

As insufficient physical activity is considered to be one of the most important risk factors for non-communicable diseases [1], physical activity should be intensively promoted in all age groups [2–4]. Due to the increasing presence of digital media in daily life, an increasing number of people worldwide is using smartphones. As a result, digital interventions in the form of applications (apps) can be developed within healthcare contexts [5]. Apps have the potential to contribute to health care [6] while remaining relatively inexpensive [7]. Recent results suggest that physical activity can be effectively increased by smartphone-based interventions [8–11].

However, an app can only be effective if it is used. The challenge lies in developing interventions for a broad target group with different expectations and lifestyles [12]. According to theoretical models of technology uptake, intention to use a technological innovation highly depends on acceptance by the individual end user [13]. Acceptance is, in itself, a function of several perceived characteristics of the technology, of which ease of use and perceived usefulness have proven to be of vital importance [14]. Ease of use, i.e. whether use of an application is difficult to learn as opposed to intuitive and therefore simple [15], is often dubbed usability [16]. According to the International Organization of Standardization
(ISO), usability is defined as the extent to which a system or product can be used in a specific context to achieve specific objectives with effectiveness, efficiency and satisfaction. Effectiveness describes how completely the objectives are achieved, efficiency, how much effort was required to achieve these objectives, and satisfaction whether or not the needs of the end users were fulfilled [17]. Research suggests users to be expecting an app which is easy to use, does not require any additional effort, provides support for problems, and is designed to be visually appealing [18].

Previous studies show how important usability testing is for successful app development [19–21]. Design influences the credibility and thereby the acceptance and effectiveness of a programme [13, 22, 23]. On the other hand, systems that are not easily and intuitively to use, because information is not effectively presented, can lead to user dissatisfaction. Consequently, the potential of interactive digital health systems is lost quickly [20]. Moreover, while components of a system might be effective in achieving the intended effect, their success is hampered by a non-satisfactory user experience. In order to prevent that analysing a systems’ usability should be an important precondition [24].

In the development of apps, the needs of potential users should be considered [16]. Iterative and user-centred development is regarded as key for identifying needs and preferences of the relevant target group [24–27], and overcoming prototype issues in early development stages [28]. For teleconsulting, Esser and Goossens suggest a framework taking into account not only the content of the application itself but also background variables of the users, such as their general attitude towards technology. For gaining information on the users’ perspective on a technology, methods such as a think-aloud based usability testing, questionnaires, or field studies can be used [24].

The aim of the described test was to assess the potential difficulties that arise when using the app "VIDEA bewegt" for the first time. The research question for this study was: What are the strengths and weaknesses of the app "VIDEA bewegt" and how can the app be improved?

**Methods**

In the following, the methodology applied to test “VIDEA bewegt” for usability is described. This test forms the basis of a study to assess the overall effectiveness of the app "VIDEA bewegt" to increase physical activity, self-efficacy, and quality of life. This study was registered with the German Clinical Trials Register (DRKS): Evaluation of an app-based activity intervention for statutory health insured people. DRKS-ID: DRKS00017392. (14 June 2019)

**Think-aloud method**

In order to evaluate usability, the think-aloud method was applied, as it is a common procedure within user-centred design processes [24]. In think-aloud tests, participants are asked to verbalise their thoughts and impressions during a certain activity. In an app-based usability test, participants are asked to
describe their experience when performing certain tasks within the app, with a special emphasis on noticeable, and disturbing elements, arising problems, or enjoyable elements. This procedure provides a deep insight into a systems’ weaknesses and strengths [29].

Pretest

At the end of December 2018, a qualitative preliminary analysis with four participants was carried out in order to gain experience with the think-aloud method and judge the applicability of the concept. From January to March 2019, the plan for the specific think-aloud tests was further revised in cooperation with a communication scientist and the "VIDEA bewegt" team.

Intervention

The app "VIDEA bewegt" (see screenshots 1 and 2), the intervention this paper focusses on, aims to sustainably increase physical activity of its users. Theoretical and practical videos provide the basis of the eight-step programme. Additional features such as documentation of physical activities and synchronising of step numbers, a user forum, and chat are included.

It is one of the first exercise apps to be officially covered by health insurances in Germany as a preventative intervention.

Screenshot 1 – Start screen

After launching the app, participants are presented a first overview of the app’s goals on five slides. By pressing the button "let’s start" ("Los geht’s") the registration process begins.

Screenshot 2 – Home

After registration, the welcome page of the app opens. At the top you can see a progress bar. The different stages are presented in the centre and the menu bar at the bottom contains the items: Stages, Focus, Activity, Video+, Exchange.

Selection of participants and study setting

As 75% of all usability problems can be identified with only four test interviews [30], we chose a sample size of ten test participants to discover as many usability problems as possible.

When selecting participants for the usability test, the aim was to cover a wide range of ages, while focusing on people aged 40 and older (see Table 1: Characteristics of participants). This is based on the
fact that older people have greater difficulty using digital media, which makes usability particularly important [31]. Plus, the risk for lifestyle-related health complications such as diabetes type II or cardiovascular diseases increases with age [32, 33]. Only participants aged older than 18 were selected. The interviews were conducted at home or at the workplace in a private, quiet, and neutral environment.

**Testing procedure**

Test-participants were asked to work their way through the structure of the app based on several pre-defined goals in 20-30 minute-interviews, describing their thoughts, impressions and problems orally. They were using the app for the first time and were asked to imagine that they had discovered it in real life. A test phone was given to the participants for the test.

The participants received both information and declaration of consent from the interviewer which they agreed to orally for data protection reasons. They were also informed that they could terminate the think-aloud test at any point in time. The interviews were conducted by a total of three different team members. Of these, two were medical students and one was a psychologist.

In order to ensure a standardised and comparable procedure, a guideline was developed for the interviewers [see Additional file 1]. In this guideline, the introduction to the study, the description of the app "VIDEA bewegt", the test procedure, questions about previous experiences with health apps as well as the expectations to app were formulated for all think-aloud sessions.

The test focused on six goals, on the basis of which the test participants got to know key components of the app. Instructions for the interviewers were attached to each goal in the guideline. Suggestions for questions prompting feedback, if none was given spontaneously, were included as well (e.g.: "What are you seeing?", "What do you notice?", "What do you want to do next?", "What problems have occurred?"). Such, test participants were motivated to continuously formulate their thoughts and utter them aloud.

The six goals tested were:

- Registration in the app
- Watching three minutes of theory video
- Watching three minutes of exercise video
- Manual entry of 5000 steps
- Sending a test message in the chat
- Starting an additional video

Following the standardised test of the app and the related think-aloud, general questions were asked about whether participants could imagine using the programme in real-life, whether their expectations
had been met and whether they missed certain app features or felt existing features to be unnecessary. In addition, age, educational level, and occupation of the participants were assessed.

In April 2019, all ten test interviews were carried out within two weeks. The transcription and analysis of the transcripts took place in April and May 2019.

**Data management**

Consent to participate was given orally and no names were used during the interview. An anonymous transcription was carried out, which did not allow identification of the participants. Only members of the research team had access to the data stored on the servers of the partaking institutions. No personal data was collected at any time during the test.

**Data analysis**

The interview was recorded directly on the test smartphone. Both an audio track and the screen of the test smartphone were captured in order to be able to visually monitor any problems arising during analysis. The complete interviews were transcribed using a web software (*otranscribe*). Subsequently, the transcripts were evaluated using qualitative content analysis according to Mayring [34]. Based on the research question, three broad categories were used to structure the deductive content analysis.

- ideas for improvement
- strengths of technology, content and design
- weaknesses of technology, content and design

The inductive formation of sub-categories, further describing each of the major categories above, was intended.

Each transcript was analysed independently by two team members. This process was supported by two experienced researchers of the Technical University of Dresden. Subsequently, the independent analyses were merged in one code system and discussed further. A codebook was created, for which the most important findings on strengths and weaknesses of the app were summarised and illustrated with examples.

**Results**

**Characteristics of participants**

For the usability test, ten smartphone users aged $\geq 18$ were selected. These users were not involved in the development of the app or study.
Three participants were 20 to 30 years, seven participants 45 to 60 years old.

Table 1: Characteristics of participants

<table>
<thead>
<tr>
<th>Test-Nr</th>
<th>Codings</th>
<th>Age</th>
<th>Gender</th>
<th>Highest educational qualification</th>
<th>Occupation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>51</td>
<td>22</td>
<td>female</td>
<td>high school degree</td>
<td>undergraduate</td>
</tr>
<tr>
<td>2</td>
<td>46</td>
<td>52</td>
<td>masculine</td>
<td>doctorate</td>
<td>graduate mathematician</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>57</td>
<td>masculine</td>
<td>university degree</td>
<td>graduate engineer</td>
</tr>
<tr>
<td>4</td>
<td>45</td>
<td>50</td>
<td>female</td>
<td>polytechnic degree</td>
<td>Nurse and music therapist</td>
</tr>
<tr>
<td>5</td>
<td>36</td>
<td>50</td>
<td>masculine</td>
<td>doctorate</td>
<td>scientific associate</td>
</tr>
<tr>
<td>6</td>
<td>52</td>
<td>47</td>
<td>masculine</td>
<td>polytechnic degree</td>
<td>religious educator</td>
</tr>
<tr>
<td>7</td>
<td>43</td>
<td>22</td>
<td>female</td>
<td>high school degree</td>
<td>undergraduate</td>
</tr>
<tr>
<td>8</td>
<td>48</td>
<td>28</td>
<td>masculine</td>
<td>bachelor’s degree</td>
<td>undergraduate</td>
</tr>
<tr>
<td>9</td>
<td>47</td>
<td>48</td>
<td>female</td>
<td>university degree</td>
<td>city administration</td>
</tr>
<tr>
<td>10</td>
<td>34</td>
<td>26</td>
<td>female</td>
<td>training qualification</td>
<td>undergraduate</td>
</tr>
</tbody>
</table>

User experiences with other health apps

Four out of ten test participants had never used a health app before. One person was currently using an app to document running sessions. Two test participants described bad experiences with health apps, as they felt restricted and patronised in their daily activities by using these apps. Two participants reported neutral experiences.

Strengths

Registration process

The overall registration process was considered not to be too long, intuitive, and easy to understand.

Font/Contrast/Operating elements

Text passages were easy to read, contrasts and colours were considered adequate.

Intuitive use of the app

Six out of ten participants perceived navigation through different app features as intuitive.

Theoretical video
Six of ten test subjects expressed themselves positively concerning the clear content and understandable message of the theoretical video. The integrated animations used to present essential information were rated positively.

**Exercise video**

The video was described as natural and understandable by test participants. The exercise coach appeared friendly and presented the exercises in a motivating way, which could be easily followed. Changes in the camera perspective were considered helpful for implementing the instructions. Four out of ten test subjects stated that the sofas in the background contributed to a relaxing atmosphere.

**Activity**

The process of entering time spent with physical activity was self-explanatory and intuitive for seven of ten test participants. The opportunity to document daily activities was evaluated positively.

**Video quality**

The quality of sound and images and the resulting intelligibility of the videos presented were generally perceived as sufficient.

**Weaknesses**

**General structure**

When assessing the design of the app, the problem mentioned most often by participants was the missing explanation and introduction to the programme. Looking at the home screen, participants could not find an explanation about structure and content of stages (see Screenshot 2). Seven of ten test subjects had difficulties to understand the structure and goals of the app when using it for the first time. The programme structure divided into stages and the interpretation of stages as course weeks was not understood by four of ten test participants.

Furthermore, there were problems with the handling of basic app components. Test participants misunderstood the structure of the menu and were unable to return to previously viewed screens.

**Registration process**

Eight of ten test participants expressed criticism at various points during the registration process. Especially the early request of personal data and the confrontation with the fee of 130€ for using the app were perceived as highly irritating (see Screenshot 3 – fee). The test participants would have preferred getting an introduction to the app before confirming the terms of use and providing personal data. The absence of contraindications must be confirmed during registration. Despite the possibility of having this
term explained, it still led to uncertainty among test participants. Furthermore, a method for synchronising
step counters had to be selected. As participants did not receive any explanatory information on how to
perform this activity until this point of the registration process, they got confused.

Screenshot 3 – Fee

The message “certified course for 130€” appears, which discourages test participants.

Theoretical video

All test participants felt disturbed by different elements of the first theoretical video. The presented expert
is perceived as overemphasised and inauthentic. His expressive gestures were considered irritating. Most
test participants reported not to get inspired by the video. Only two out of ten participants felt motivated
by the video.

Exercise video

The missing introduction of the expert and the length of the video of more than 20 minutes were criticised
in this video. Furthermore, four of ten test participants worried whether it would be possible to perform the
exercises presented in the videos using only small smartphone screens as a source.

Presentation of the experts

The participants noticed that titles of the experts were not equally formulated, even though both experts
had scientific backgrounds.

Documentation of activity

Nine out of ten test participants expressed themselves critically during the task of entering an activity. An
explanation of the objective and need for documentation of activity was missing. Furthermore, the list of
everyday life activities to select from was not exhaustive. The origin of the indicated activity goal was not
understood at first sight. The app feature of tracking one's mood raised the question of why and when
this should be done. Eight out of ten test participants evaluated the design and structure of the feature for
tracking activities as negative. The feature for entering step counts was difficult to find and to
understand. Operating the various elements and understanding the texts was challenging (See
Screenshot 4 – Activity). For example, test participants had difficulties to enter the correct date while
some participants wondered whether values, once entered, could be corrected later. Frequent freezing of
the app and delays when operating the menu led to immediate frustration of test participants.

Screenshot 4 – Activity

Activity is illustrated by a circle that represents the progress made towards the daily goal of activity
minutes. Below participants will find a short motivational message, a statistic of the number of steps
achieved and their course of the mood over the last week.
Expert chat

The expert chat was not immediately found within the app by three test subjects and sending a message in the chat was not perceived as intuitive.

Slowness of the App

Nine out of ten test participants reported that the app did not run smoothly. The time the app needed to start was very long and was interrupted by crashes. Control elements did not react immediately several times, which led to uncertainty and frustration among test participants. These problems of response time mainly occurred while entering minutes of activity and watching videos.

Further problems

During the tests a number of minor errors were noticed, which further complicated the app use. For example, the welcome message contained a text ending in a blank. Some videos were linked incorrectly, which led to confusion when trying to watch a specific video. It was also noticed that the app is not perfectly adapted to every available smartphone screen size. As a result, test participants with smaller smartphones might need to scroll to obtain essential buttons and information.

As a general observation, the question whether a permanent internet connection was needed for app use was raised.

Suggestions for improvement

The main suggestions made by the participants were:

- At the beginning of the programme an introduction to the structure and functions of the app would be helpful.
- A pedometer integrated in the app itself would improve usability.
- It would be more motivating to formulate personal goals instead of only being asked about the current activity level.
- A direct feedback to the registered activity would be helpful for assessing personal performance.
- The choice of activities should be either more detailed and include all possible sports and activities or be more general.
- An individually adjustable font size would be useful as the current font size is well suited for young users but may be too small for older people.

General Feedback of the participants
All in all, the most important expected features were found, but the implementation and user-friendliness were criticised. After the test, four of ten participants could imagine using the app in the future, one person was not sure and five considered it to be unlikely.

Discussion

In the test described, the video content and other app-integrated features were rated positively. The app was appreciated for its design, basic registration process and the largely self-explanatory navigation. Colours, contrasts, and fonts were predominantly judged positively. The early retrieval of personal data and the financial fee presented in the registration process were commented on negatively by test participants. Paywalls are a well-known problem when distributing health information online, as they discourage users to further use an application [35].

Users missed an explanation of the included components and the basic structure of the programme, a need that can be traced back to its theoretical roots in facilitating conditions. Those can, according to a current review on technology acceptance, facilitate the uptake of a certain health application. Barriers for acceptance have been shown before, such as users not understanding the scope, functionalities and menus of apps [36]. Due to a lacking introduction to the app, documenting activities was not recognised as an essential part of the programme. Entering activities caused problems with control panels not being self-explanatory. This is in line with the results of Georgsson’s and Staggers’ analysis of diabetes mHealth applications [37].

The importance of a user-centered design process is demonstrated once more, as it makes sure that an application is indeed used the way it was intended by its developers [27, 38]. Test participants also suggested the app content should be tailored directly to the behavioral parameters they reported, which is in line with existing research on measures to improve usability [39]. Also, smartphone applications generally allow for tailoring of content [40] and individual goal setting [41], which was also wished for by the participants.

Knowing that multifunctional apps are often less user-friendly than simpler apps, and that potential users should be involved in the design and development of a product in order to ensure its success [42], all problems and suggestions for improvement were forwarded to the institution who commissioned the app for further revision.

As intended by the circular process of user-centered design [43], first corrections to design and structure of “VIDEA bewegt” have already been made. The biggest and most important change has been an introductory video that explains the structure and features of the app and the programme flow. Plus, the presentation of the discouraging fee for registration has been modified in the registration process. Delays and lags have been removed as far as possible. In addition, some of the problems encountered during first use are likely to resolve quickly after a short period of learning or adaptation, which is in line with the diffusion of innovation process described by Rogers [44].
Strengths and Limitations:

Think-aloud is an established method for understanding the thought processes and problems of test subjects [20, 24]. The advantages of this method are that information can be collected continuously without a large number of specific questions. The test can help identifying problems and difficulties at a very early development stage. With people communicating their thoughts orally, a maximum amount of information can be collected ad hoc while the users are familiarising themselves with the new tool [30]. As a certain subjectivity is common to self-reported data, combining think-aloud with observational results has become standard [20]. Such, the think-aloud sessions for “VIDEA bewegt” were videotaped.

One of the great strengths of the think-aloud method is the potential to identify 75% of all usability problems of a system with only four test interviews. Therefore, very few test subjects are needed to gain important insights [30]. However, although the sample size of ten people is already comparatively large for a think-aloud test [45], the results are based on subjective impressions and can only be quantified to a certain, rather descriptive degree. Furthermore, the test participants are only representative of the target group to a limited extent. For example, all test subjects had academic backgrounds and used digital media to a varying extent. The age structure was intentionally focused on persons older than 40 years, because it is expected that persons in this age group represent the most important target group of the app as a preventative intervention. However, no consideration was given to whether test persons were actually interested in using an app such as "VIDEA bewegt". Plus, the test covers only first impressions of app usage and does not allow for tracking changes in perceived ease of use of a certain time of use.

As for the think-aloud method, it can be one disadvantage that people perceive the method as disturbing. In addition, some people might struggle to express their thoughts precisely [29].

Test sessions were not conducted with the test subjects' phones, but with test devices, which in some cases differed from the handling of their own phones. The test interviews were conducted by three interviewers. For this reason, the interview structure differed slightly regardless of the interview guidelines. This procedure, however, limits any potential interviewer bias. Conducting the analysis of the transcripts and videos with two independent researchers served the same goal.

Future research

It remains an open question how quickly test persons would adapt to the structures and features of the app and how satisfied they would be with its use over a longer period of time. These questions are intended to be answered by a process evaluation within the framework of the study. In order to recognise an improvement in usability, usability tests should be repeated after elimination of problems [27].
**Conclusions**

A major weakness of the apps’ design results from test participants not understanding central elements and basic structures. From a technical point of view, both picture and sound quality of the videos are consistently good. However, long loading times and stalling of the videos need to be improved. Basically, the ten test interviews showed that all test subjects were able to use the app independently to a large extent.

Looking at the definition of usability as a combination of effectiveness, efficiency and satisfaction, most of the a priori defined goals were achieved successfully. The effectiveness can therefore be assessed as good. In summary, it is reasonable to assume that the app has the potential to offer a well-functioning programme.

**Declarations**

**Ethics approval and consent to participate and for publication**

After consultation of the Ethics Committee of the Technical University of Dresden, no extensive ethics vote was necessary for this test, as long as no person-related data was collected.

The informed consent (according to the specifications of the Ethics Committee) to participate in this test was given orally.

**Consent for publication**

Not applicable

**Availability of data and materials**

The datasets generated and/or analysed during the current study are not publicly available due to data protection reasons but are available from the corresponding author on reasonable request.

**Competing interests**

Prof Schwarz was involved in the development and implementation of the app "VIDEA bewegt" as a medical expert. He is responsible for the medical and theoretical background and is shown in the app's videos.

He received no payment for his participation in the app.

**Funding**

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.
We acknowledge support by the Open Access Publication Funds of the SLUB/TU Dresden.

Authors' contributions

TF, GR and PS conceived and designed the test, wrote the manuscript, collected the data, and performed the data analysis.

PEHS provided feedback on the overall procedure and reviewed the manuscript.

LH regularly provided feedback on the procedure, was involved in the development of the usability test, and participated in writing the manuscript.

All authors reviewed and approved the final version of the manuscript before submission.

Acknowledgements

We would like to thank Sarah B. Veith for her support with the language revision.

References


42. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: A systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. J Med Internet Res. 2014;16.


Additional File Legend

Additional file 1: docx format, “Guideline for interviewers”, This document contains a translation of the guideline, which was used as a standardised basis for carrying out the tests.

Figures
Begeben Sie sich auf Ihre Reise zu Ihrem aktiven Selbst!

LOSERGEHT’S

Ich habe schon ein Konto

**Figure 1**

Start screen
0 % abgeschlossen
Sie haben 0 von 8 Etappen abgeschlossen

Etappe 1
Bewegung liegt uns in den Genen
Sie erfahren, warum Bewegung so wichtig ist, werden gemeinsam mit Ivonne aktiv und stärken Ihre Motivation.

STARTEN

Figure 2

Home
Zertifizierter Kurs für 130,-€

Ihre Krankenkasse erstattet Ihnen bis zu 100% der Kursgebühr oder zahlt sogar vorab.

Zum Erstattungsrechner

Sie haben schon ein Konto?
Zur Anmeldung

Figure 3

Fee
**Figure 4**

Activity

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- Guidelineforinterviewers.docx