Original Paper

Is a mHealth application for pregnancy related work advice usable and useful?

A mixed method approach.

Authors:
M.D.M. van Beukering
A.V. Velu
L.A. van den Berg
M. Kok
B.W.J. Mol
M.H.W. Frings-Dresen
R.A de Leeuw
J.A.M. van der Post
L.W. Peute

Affiliations:
1 Department of Obstetrics and Gynecology, Amsterdam University Medical Centers, location AMC, Amsterdam, The Netherlands
2 Department of Medical Informatics, Amsterdam University Medical Centers, location AMC, Amsterdam, The Netherlands
3 Monash University, Melbourne, Australia
4 Department Coronel Institute of Occupational Health, Public Health research institute, Amsterdam University Medical Centers, location AMC, Amsterdam, The Netherlands
5 Department of Obstetrics and Gynecology, Amsterdam University Medical Centers, location VUMC, Amsterdam, The Netherlands

Corresponding author:
M.D.M. van Beukering
Department of Obstetrics and Gynecology
Amsterdam University Medical Centers, location AMC
P.O. Box 22660
Amsterdam, 1100 DD
The Netherlands
m.d.vanbeukering@amc.uva.nl
Keywords:
mHealth; Pregnancy; Work; Occupation; Exposure; Think Aloud, Mixed Methods
ABSTRACT

**Background:** Pregnant women are often unaware of potential risks for mother and child due to working conditions. A mHealth application, the 'Pregnancy and Work' app, developed by a multidisciplinary team and based on an evidence-based guideline for occupational physicians, aims to provide advice on work adjustment during pregnancy.

**Objective:** This study evaluates the usability of the mHealth 'Pregnancy and Work' app in relation to the perceived usefulness of the given work advice (the main goal of the app) by potential end-users.

**Methods:** Potential end-users (n = 12) participated in Think Aloud sessions and performed nine tasks. All Think Aloud sessions were recorded, transcribed and co-analysed. The usability problems were rated for their severity in accordance to Nielsen’s severity scale. The completion rates and times of tasks were registered. Additionally, participants were questioned on demographics, user characteristics, value and evaluation of the app by the Intrinsic Motivation Inventor and perceived usability by the System Usability Scale.

**Results:** Twelve participants identified 82 usability problems with a severity ≥ one, of which 40 with severity ≥ three. Main usability problems concerned interpretation of terminology used in the app’s questionnaires and difficulties in finding and understanding the work advice. Ten out of 12 participants were able to open the work advice page in the app. Only seven out of these ten participants understood and intended to use the work advice. The overall mean Intrinsic Motivation Inventory score was relatively high (five out of seven) indicating that participants indeed valued the use of the app. This Intrinsic Motivation Inventory score corresponded to the overall mean System Usability Scale score (68 out of 100) and the mean grade given to the ‘Pregnancy and Work’ app (seven out of ten).

**Conclusion:** This Think Aloud usability study showed that the information provided in the 'Pregnancy and Work' app was considered valuable to the end-users and meets their need, but usability issues severely impacted the perceived usefulness of the work advice given in the app.
1. Introduction

Many women continue to work during their pregnancy. In the US, more than 65% of pregnant women work, while in the Netherlands this is 80% [1, 2]. During pregnancy, exposure to certain working conditions, like physically demanding work, long working hours, working night shifts and stress are associated with preterm birth, low birthweight and fetal abnormalities [3-12]. Because pregnant women are often not aware of these risks, they do not adjust their working conditions [13]. Mobile Health (mHealth) applications can form a suitable solution to this problem, as women of reproductive age that are expecting a child are frequent consumers of online health information [14-17]. MHealth, defined as the use of mobile devices for medical and public health practice [18], could therefore inform pregnant working women about work-related pregnancy risks to increase their awareness of these risks and their associated need for change in working conditions. However, current research shows that mHealth apps in general continue to proliferate with little evidence for their effectiveness [19] and little support for understanding how best to design them [20-24].

In this study we developed a mHealth solution that aims to provide information and advice about work-related pregnancy risks, the 'Pregnancy and Work' app (P & W app) [25]. The P & W app’s content is based on the evidence-based guideline for occupational physicians: Pregnancy, Postpartum Period and Work [26]. In designing mHealth adequate consideration of the needs of their intended users is necessary so that they are easy to use and perceived as useful [27,28]. In a prior study the results of two multidisciplinary focus group meetings provided content and design instructions for the development of the P & W app [29]. The research addressed in this paper focuses on the evaluation of the user interface design and user testing of the application: testing the application via a mixed-method approach on its usability and perceived usefulness. The applied definition for usability in this study is based on the International Organization Standardization (ISO): “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [30]. Usability evaluation methods (UEMs) assess whether users value, understand and act on the advice given. UEMs can be utilized to assess the human interaction with a system for the purpose of identifying facets of this interaction that can be improved [32,32].

Typically, a design process of any health-related product is conducted in an iterative fashion. Utilizing UEMs in such an iterative design process in the health care domain is especially important as poor design and usability of medical products can lead to harmful consequences [33,34]. Therefore, the utilization of UEMs during the development and testing process is widely recommended throughout research [35,36]. The Think Aloud (TA), a UEM method, was chosen in this study, to assess the usability of the P & W app with their potential end-users to reveal their cognitive processes in interaction with this app that result in user-interaction problems [37]. The TA method requires participants to talk aloud (i.e., verbalize their thoughts) while performing or solving a task [38-40].
This study will take an in-depth look at the evaluated usability and experienced usefulness by potential end-users. The study aims to evaluate the usability of the mHealth ‘Pregnancy and Work’ app in relation to perceived usefulness of the work advice given in the app.

2. Methods

Study set up: a mixed methods design (case study).

Procedure

Two obstetric care facilities representing a broad variety of patient groups, participated in this study. Posters and flyers were distributed on both locations. The inclusion criteria were: Dutch working women, whom were less than 20 weeks pregnant. Eligible participants were recruited in the waiting area. Recruitment of participants continued until a total of 12 female patients evaluated the app and would partake in the TA sessions, it was the first time they used the P & W app. All participants that were included in the study would receive a gift card worth €15.

P & W Application under study

The P & W app (Dutch and English) is created as a web application, which means that it is accessible from every mobile devices browser, because of costs consideration and the availability from desktop. The P & W app requires the user to create an account to gain access to its content. After creating an account, users have to fill in a questionnaire about their (pregnancy related) medical- and work conditions (Figure 1). When completing this questionnaire, the user will be directed to the home page of the app from this home page the user can navigate to all pages of the app. On the home page users can view monthly pregnancy and work-related advice messages, which are also sent by e-mail. The application furthermore provides baby messages that entail information about the growth of the unborn baby as the weeks pass by. Next to this, a video, with tips and information about pregnancy-related work advice can be viewed on the home page.

Figure 1: Examples of screenshots of Pregnancy and Work App

Welcome page, Questionnaire page, Workadvice page
A Dutch beta-test version of the P & W app applied in this study was a webpage adjusted for mobile phone utilization. This allowed for easy access to the P & W app for the participants, regardless of their phone type or operating system. Thus, participants were first time users. After recruitment arrangements would be made to set up the TA session, after their next consultation at the obstetrics department. The following study set-up for the usability testing was composed of three phases (figure 2). In the next sections, each phase is described in more detail.

Figure 2. Overview of study setup.

Stage I: Preparation
In the first phase of the study set-up the participants were informed about how the TA session would be performed. With participants who wanted to participate in the study, an appointment was made after two weeks reflection period. After filling in an informed consent form the participant would fill in a short survey to assess their level of health literacy (Stage I). The Newest Vital Sign, a validated method translated in Dutch (NVS-D), was utilized to assess the participant’s health literacy (HL) level [41,42] (Appendix II).

Stage II: Think Aloud Usability testing experiment
In stage II the participant started with a practice tasks on how to ‘think aloud’. The participant was informed that the researcher (LvdB) was solely interested in app performance and would only interrupt the participant to provide new tasks and to encourage her to keep talking in order to break silences longer than five seconds [37,43]. A participant had to complete nine tasks in total that were centered around the core purpose of the app (Appendix III). Tasks were developed in collaboration with the developer and project supervisors of the P & W project. All TA sessions were recorded via means of a video camera. Voice and screen (of their mobile phone) were also recorded. (Figure 3)
Stage III: Questionnaires

After the TA test was finished. The System Usability Scale survey was given to the participant to assess the perceived usability of the P & W app [44] (Appendix IV-1). The System Usability Scale consists of 10 statements to which the participant must rate on a scale from one (strongly disagree) to five (strongly agree) as to which extent they agree with that statement. Then, a short survey selection of the Intrinsic Motivation Inventory was given to assess a self-reported evaluation on how much the participant value the P & W app [45] (Appendix IV-II). The Intrinsic Motivation Inventory value subscale consist of seven statements where the participant has to rate on a scale from one (do not agree) to seven (strongly agree) to which extent they agree with that statement. An additional short survey was developed to gain more insight into participants’ understanding of the work advice given in the app, their intent to follow-up on the advice given, their demographics, medical history related to pregnancy, prior experience with (pregnancy related) mobile applications and working hours (Appendix IV-III). Lastly, participants were tasked to give the P & W app a grade on a scale from one to 10, where one was the lowest and 10 the highest grade.

Data collection and analysis

TA sessions were videotaped, reviewed multiple times and transcribed to verbal protocols by two researchers (LvdB, LP). To gain insight into the effectiveness and efficiency of the participants in performing tasks, each TA session transcription was comprised of text spoken by the participant and included task completion time stamps and task completion times. To analyze usability problems in detail a coding scheme was developed bottom-up, based on two transcribed verbalizations. The developed coding scheme was subsequently applied to code and analyze all verbalizations, by LvdB and checked by LP. Usability problems were rated on severity in accordance to Nielsen’s severity classification [46]. Nielsen’s severity scale is a rating scale from zero to four (figure 4) that allows for prioritization of usability problems that need to be revised in the development process. The questionnaires were filled in on paper and put in a .csv file for data analysis.

Figure 4. Nielsen’s severity scale.
0 – I do not agree that this is a usability problem at all.
1 – Cosmetic problem only: need not to be fixed unless extra time is available on project.
2 – Minor usability problem: fixing this should be given low priority.
3 – Major usability problem: important to fix, so should be given high priority.
4 – Usability catastrophe: imperative to fix this before product can be released.

All data that was filled in by participants in the P & W app during the TA sessions were specifically transcribed in a different file to test for task efficacy in relation to the Intrinsic Motivation Inventory given work advice by the system, the second aim of this study. Verbalizations of task six in the TA sessions (‘find the work advice’) were assessed to analyze whether participant would follow up the work advice. These results were compared to the results of the Motivation Inventory on participant level and questions about the work advice from the questionnaire III (Appendix IV-III). Finally, the System Usability Scale was used to assess the perceived usability of the P & W app.

3. Results
3.1 Participants characteristics
The TA sessions with the participants (N=12) took place in April until June 2017. Most participants scored a high (= adequate) health literacy. All participants had paid jobs, and a used smartphone. The average gestational age of the participants was 15 weeks, 50% of the participants were pregnant for the first time (table 1).

Table 1. Participant basic demographics and characteristics.

<table>
<thead>
<tr>
<th>Participants</th>
<th>N= 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (years)</td>
<td>33 (SD = 3.8)</td>
</tr>
<tr>
<td>Education: secondary school</td>
<td>100% (N=12)</td>
</tr>
<tr>
<td>Higher education = 8, Intermediate vocational education = 4</td>
<td></td>
</tr>
<tr>
<td>Health Literacy</td>
<td>High: 92%</td>
</tr>
<tr>
<td>Low: 9%</td>
<td></td>
</tr>
<tr>
<td>Paid job</td>
<td>100% (N = 12)</td>
</tr>
<tr>
<td>Working time (hours a week)</td>
<td>37 (SD = 6.15)</td>
</tr>
<tr>
<td>Gestational age (weeks)</td>
<td>15 (SD = 3)</td>
</tr>
<tr>
<td>Previous pregnancy</td>
<td>50% (N = 6)</td>
</tr>
<tr>
<td>Kids</td>
<td>42% (N = 5)</td>
</tr>
<tr>
<td>Smartphone (Operating System)</td>
<td>100% (N = 12)</td>
</tr>
<tr>
<td>Android: 7</td>
<td></td>
</tr>
<tr>
<td>iPhone: 5</td>
<td></td>
</tr>
</tbody>
</table>

3.2 Task completion
The results for the first aim of this research, the effectiveness and efficiency of the participants in performing tasks, concern the completion rates and times and usability problems. The completion rates and times can be found in table 2. The average duration of a TA session was 19:55 minutes (SD = 05:25). Task two, three, five and nine were completed by all participants. Task one, four, six, seven
and 8 were not. The first three tasks took, on average, the longest to complete ranging from four minutes to one minute and a half. Task nine had the fastest mean completion rate of four seconds.

Table 2. Completion rates and times per task (9) participants performed.

<table>
<thead>
<tr>
<th>Task</th>
<th>Completion rate</th>
<th>Mean completion time seconds (min)</th>
<th>SD (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Create an account.</td>
<td>10/12</td>
<td>240 s (4:00)</td>
<td>(SD = 83)</td>
</tr>
<tr>
<td>2. Fill in questionnaire.</td>
<td>12/12</td>
<td>179 s (2:59)</td>
<td>(SD = 101)</td>
</tr>
<tr>
<td>3. Adjust answers questionnaire.</td>
<td>12/12</td>
<td>96 s (1:36)</td>
<td>(SD = 74)*</td>
</tr>
<tr>
<td>4. Find 'Your rights and tips for consultation' page</td>
<td>11/12</td>
<td>31 s</td>
<td>(SD = 38)</td>
</tr>
<tr>
<td>5. Find 'baby message(s)'.</td>
<td>12/12</td>
<td>16 s</td>
<td>(SD = 10)</td>
</tr>
<tr>
<td>6. Find the 'Your work advice' page.</td>
<td>10/12</td>
<td>10 s</td>
<td>(SD = 8)</td>
</tr>
<tr>
<td>7. Find the 'PRINT/SAVE' button.</td>
<td>10/12</td>
<td>9 s</td>
<td>(SD = 9)</td>
</tr>
<tr>
<td>8. Find the goal of the P &amp; W app.</td>
<td>11/12</td>
<td>32 s</td>
<td>(SD = 18)</td>
</tr>
<tr>
<td>9. Log out of the app.</td>
<td>12/12</td>
<td>4 s</td>
<td>(SD = 4)</td>
</tr>
</tbody>
</table>

* Two participants initially did not understand this task.

3.3 Usability problems

The Think Aloud study identified a total of 101 usability issues, 82 of which were considered ‘real’ usability problems (i.e., severity ≥1), 40 usability problems were rated with a severity of three (major) or four (catastrophic). Additionally, the participants encountered 11 unique bugs when using the P & W app. An overview of the most severe usability problems can be found in Table 3. None of the participants experienced (severe) usability problems when completing task five, seven and nine. In the following section we give an in-depth analysis of the severe usability problems detected regarding ‘Terminology interpretation’ and ‘interpreting and understanding the workadvice’ that directly impacted the participants perceived usefulness of the advice given in the app.

Table 3. Overview severe usability problems per main problem type*.

<table>
<thead>
<tr>
<th>Usability problem</th>
<th>Frequency</th>
<th>Severity</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unclear buttons</td>
<td>12</td>
<td>2-4</td>
<td>Create account</td>
</tr>
<tr>
<td>Functionality with lay out</td>
<td>11</td>
<td>4</td>
<td>Create account/ Home page</td>
</tr>
<tr>
<td>Terminology interpretation problems</td>
<td>8</td>
<td>4</td>
<td>Create account/ Home page</td>
</tr>
<tr>
<td>Interpreting and understanding workadvice</td>
<td>8</td>
<td>4</td>
<td>Home page/ Work advice</td>
</tr>
</tbody>
</table>

*An overview of the entire usability problems is requestable.

3.4 Qualitative assessment

Terminology interpretation problems

Participants had to fill in a questionnaire about their pregnancy related medical conditions, previous pregnancy (if relevant) and their work conditions in the app. Several terminology interpretation problems did arise during the TA study which consequently inhibited the participants in receiving an accurate personal work advice. For example, when asked whether during the previous pregnancy...
problems had been experienced, participants were unsure whether ‘previous pregnancy’ implied the
direct previous pregnancy, or also those before. One participant who had not experienced problems
during her previous pregnancy, but did experience issues the pregnancy before that one, assumed it
implied her direct previous pregnancy. Her confusion in answering the question correctly effected
the outcome of the work advice due to missing relevant information.

“Okay. Um ‘Did you have a medical problem in your previous pregnancy?’ This is about my last
pregnancy, I think, and not the pregnancies before. So, I’m assuming that. And then it’s a no.”
-participant – 5

Problems also were prevalent when closed ended questions were stated, within the limited selection
of possibilities of medical disorders the participant did not find the answer that applied to her. When
given a list of potential problems to select of that were encountered during a previous pregnancy,
participants experienced troubles in selecting the best suited option to describe their problem.

“… But I do not know if that can put under ‘deceased child’ or ‘child born before a gestational age
of 37 weeks’? You know what I mean?” – participant 3

Another example of a terminology interpretation problem that affected the outcome of the working
advice related to the question of ‘being exposed to any chemical agents in the work environment’,
followed by a list of examples. Several participants did not notice the list of examples and answered
‘no’. Two other participants did not know whether an agent that they worked with should be
considered chemical as it was not on the list of examples.

“… Yes, with hair dye. Is that chemical?” – participant 9

“… I’m having doubts. I work with laughing gas. That’s not very chemical, but… I don’t know
whether I should answer yes or no.” – participant 11

Interpreting and Understanding the work advice

Participants also experienced problems with understanding the work advice because of central design
problems in the interface. One of the first issues encountered was that participants had a different
expectation of what the app showed them. Their mental model, the way information is represented
in the mind of the end-user, in filtering relevant information, affected how they acted in the system.
This did not match how the designer developed the system based on his mental model of how future
end-users should act on information presented. Participants expected the app to show their work
advice directly on the homepage, as they perceived it as the essential goal of the application. They
did not expect they had to search for it in the interface or take any other action to find the work
advice. For example, participant 6 did not understand that the [Your work advice] button was
clickable and therefore sought work advice elsewhere or stated that they could not find it (figure 5).
“Oh, let’s see if that is somewhere. No idea.” Scrolls down and up. “Have a look. Here is my work advice. Uh ... “Scrolls up and down, multiple times. No, I have no idea.” – Participant 6

A different example in user mental model system mismatch related to participants stating that they actually saw their work advice depicted on the home page. However, the home page only provides a small section with tips and information about pregnancy-related work advice. To some this was interpreted as the entire personal work advice. Two participants thought this was the case and therefore both missed the actual content of the ‘Your work advice’ page (see participant 8).

“None? That’s easy. I don’t need to make any work adjustments. I don’t think so either, because I have an office job.” – participant 1

A different problem related to the inconsistency in the work advice in providing information. Depending on which answers were given in the questionnaire, specific information followed on the work advice page. The resulting advice therefore included sections having no advice, and some do, all spread over the mobile interface. One participant did not get a work advice below the ‘work header’, however she did receive work advice with regard to issues during her previous pregnancy. She therefore missed the advice given.

“None? That’s easy. I don’t need to make any work adjustments. I don’t think so either, because I have an office job.” – participant 1

**User evaluation: Intrinsic Motivation Inventory and System Usability Scale**

The task efficacy of task six ‘Find the ‘Your work advice’ page’ was analyzed in relation to the detected usability problems in ‘interpreting and understanding the work advice’ and combined with the results of the Intrinsic Motivation Inventory, the System Usability Scale and questions about the work advice from the questionnaire III (Appendix IV-III). Some participants never reached the work advice page on the app (17%) but thought they did, three out of 12 participants (25%) were convinced they had not
received such advice (table 4). However, all participants actually did receive some form of pregnancy-related work advice. Among the nine participants that did state that they received work advice, two indicated that they would not utilize the advice.

Table 4. User evaluation:

(Use the) work advice, Intrinsic Motivation Inventory (IMI), System Usability Scale (SUS), Grade

<table>
<thead>
<tr>
<th>Work advice?</th>
<th>Use advice?</th>
<th>IMI</th>
<th>SUS</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No</td>
<td>NA</td>
<td>5.57</td>
<td>85</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>NA</td>
<td>4.29</td>
<td>77.5</td>
</tr>
<tr>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>3.71</td>
<td>55</td>
</tr>
<tr>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
<td>5.00</td>
<td>77.5</td>
</tr>
<tr>
<td>5</td>
<td>Yes</td>
<td>Yes</td>
<td>5.14</td>
<td>65</td>
</tr>
<tr>
<td>6</td>
<td>Yes</td>
<td>No</td>
<td>4.43</td>
<td>77.5</td>
</tr>
<tr>
<td>7</td>
<td>Yes</td>
<td>Yes</td>
<td>5.57</td>
<td>57.5</td>
</tr>
<tr>
<td>8</td>
<td>Yes</td>
<td>No</td>
<td>3.00</td>
<td>70</td>
</tr>
<tr>
<td>9</td>
<td>No</td>
<td>NA</td>
<td>4.29</td>
<td>75</td>
</tr>
<tr>
<td>10</td>
<td>Yes</td>
<td>Yes</td>
<td>6.29</td>
<td>55</td>
</tr>
<tr>
<td>11</td>
<td>Yes</td>
<td>Yes</td>
<td>5.29</td>
<td>50</td>
</tr>
<tr>
<td>12</td>
<td>Yes</td>
<td>Yes</td>
<td>4.86</td>
<td>72.5</td>
</tr>
</tbody>
</table>

*NA = not applicable.

Via the Intrinsic Motivation Inventory we assessed the self-reported evaluation on how much the participants valued the P & W app: the overall mean Intrinsic Motivation Inventory value score was a five (SD = 0.9) out of seven. The perceived usability of the P & W app was stated by the System Usability Scale. The overall mean System Usability Scale was 68 (SD = 11). Lastly, participants were tasked to give the P & W app a grade on a scale from one to ten: the mean grade given to the P & W app was a seven (SD = 0.89) (table 4) Reliable correlation calculations between the completion times, System Usability Scale, Intrinsic Motivation Inventory, and grades given by the participants could not be assessed due to sample size.

4. Discussion

Principal Findings

Overall effectiveness and efficiency of the twelve participants in performing tasks in the TA sessions concern the completion times and rates and usability problems. Main usability problems concerned interpretation of terminology used in the app’s questionnaires and difficulties in finding and understanding the work advice. The Think Aloud study identified 82 usability problems with a severity ≥ one, of which 40 with severity ≥ three. We performed an in-depth analysis of the severe usability problems detected regarding ‘terminology interpretation’ and ‘interpreting and understanding the work advice’ as these issues directly impacted the usefulness of the app. Because participants were unable to correctly interpreted the terminology in the questionnaire about previous pregnancies, medical disorders and chemical agents they did not understand how to correctly fill in the
questionnaires so that it corresponds to their personal situation. They thus could not act accordingly and did not receive the correct personal work advice. Participants also had a different expectation of what the app showed them. Their mental model, which encompasses values, beliefs, and knowledge that create perspectives for filtering information and guiding problem solving [47], and has the ability to affect how a person acts [48], differed from the designers and they therefore also experienced problems with understanding the work advice.

Due to usability problems in its design, only seven out of 12 participants (58%) were able to get to and/or understand and had the intention to use the work advice given in the app: the main goal of the app.

The overall mean Intrinsic Motivation Inventory value score was relatively high (five out of seven) indicating that participants indeed valued the use of the app. This corresponded to the overall mean System Usability Scale (score 68 out of 100) and the mean grade given to the P & W app (seven out of ten).

Comparison with prior work

Studies in mHealth and eHealth that have applied the TA method have demonstrated, though participants ‘think’ they have achieved the main goal of using the applications, in truth they did not reach it’s intended effect. This was also the case in our study. In one study the researchers observed that the large majority of participants, older cancer patients, were not able to find requested information. Participants however frequently commented during testing that it was easy for them to find it [49]. In a different study, patients with rheumatic diseases were enthusiastic about possibilities of interactive applications such as peer support forums and online consults, however, nearly all participants experienced difficulties and were not able to complete all of the usability evaluation tasks while interacting with the system [50].

Other studies indicated to the importance of using text message or interactive and individual coaching to improve the lifestyle of pregnant women (nutrition, smoking, weight, alcohol) [14-16]. When using the P & W app, the aim is to provide work advice so that pregnant woman will act upon the personal work advice given. In understanding that the app design influences the understanding and receiving of the work advice, we therefore aim to regularly send emails during the pregnancy with an updated work advice. A pregnant employee can then discuss this with their supervisor, occupational physician, midwife or gynecologist. This interaction with healthcare providers can increase the effect of the P & W app.

Similar to our study, other researchers and designers also underline the importance of an iterative approach, to design mHealth applications with understanding the needs of end users as well as how to improve app usability and feasibility [36,51]. The importance of performing usability studies on
mHealth applications to be used in a clinical and/or patient setting therefore needs serious attention. User testing is an essential part of developing mHealth applications, especially when aiming to actually effectively change patient behavior and/or affect patient outcomes.

**Strengths and limitations**

A limitation is that the Think Aloud sessions took place in a laboratory setting. In their own home, participants might have taken more time to take a look at the app again. Strengths are that the sample size in this study is adequate for obtaining usability problems and that we used a mixed methods approach: we combined the results of a Think Aloud test with the results of questionnaires on demographics, user characteristics, System Usability Scale, perceived value (Intrinsic Motivation Inventory) and evaluation of the app. Another strength of our study is that it was performed in a multidisciplinary team and that the Think Aloud study is part of a process in developing a mHealth application, which started with two multidisciplinary focus group meetings [29].

The mixed method approach provides insight into the cognitive process of participants and their intention to use the P & W app. The Think Aloud results in combination with the questionnaires on perceived usability and value and evaluation of the app showed that, though participants had an adequate health literacy, incorrect interpretation of terminologies in the system prevented the end users to get the correct work advice. They also experienced problems with understanding the work advice because of central design problems in the interface. Despite many usability problems, the participants were relatively positive about the P & W app, information provided in the app is considered valuable to the end-users and meets their need.

**Conclusion**

The overall conclusion of this study is that the information provided in the P & W app was considered valuable to the end-users and meets their need, but usability issues severely impacted the perceived usefulness of the work advice given in the app. The results of this study will be integrated in redesign of the app and tested in a intervention study, a survey on the effect of the app on actual work adjustment by pregnant women. Continuous research, user testing, and improvements are facets that need to be addressed in order to realize a future version of the P & W app as an appealing and valuable tool for informing pregnant women about pregnancy-related work risks.

**Ethical approval**

An application for this research has been submitted to the ethical board of the Amsterdam University Medical Centers, location AMC. The board confirmed that the Medical Research-Involved Human Subjects did not apply to this study. All data from the twelve participants were anonymously
processed. Informed consent was obtained from all participants, allowing us to use the data for analysis.

Acknowledgements
We would like to thank all the pregnant participants who participated in the Think Aloud sessions and the employees of the obstetric care facilities.

Conflicts of interest
All authors declare that they have no conflicts of interest.

Funding
This pilot study had received funding from ZonMW, the Netherlands Organization for Health Research and Development. This project is part of the Pregnancy and Birth program.

Abbreviations
mHealth= Mobile Health
P & W app= Pregnancy and Work app
UEMs= Usability Evaluation Methods
TA= Think Aloud
References


