Enhancing the Usability of a Mobile App for Process Evaluation in a Participatory Ergonomics Healthcare Intervention

Winnie Chin[1,2(✉)], Alicia Kurowski[1,3], Guanling Chen[4], Rebecca Gore[1,3], and Laura Punnett[1,3]

[1] Center for the Promotion of Health in the New England Workplace (CPH-NEW), Lowell, MA 01854, USA
winnie_chin@student.uml.edu
[2] Department of Public Health, University of Massachusetts Lowell, Lowell, MA 01854, USA
[3] Department of Biomedical Engineering, University of Massachusetts Lowell, Lowell, MA 01854, USA
[4] Department of Computer Science, University of Massachusetts Lowell, Lowell, MA 01854, USA

Abstract. Monitoring of workplace intervention processes in real-time can identify factors influencing intervention success or failure while they can still be modified if necessary. The aims of this study are to describe a process to assess and (where necessary) improve functionality and usability of a process evaluation mobile app. The app was developed for the research study "Safety and Health through Integrated, Facilitated Teams" (SHIFT), an evaluation of the CPH-NEW Healthy Workplace Participatory Program in public sector healthcare institutions. App users are members and co-facilitators of labor-management health and safety committees, internal program champions, other managers, and researchers. The app records four "functions:" meetings held, attendance and attendee feedback; project-related chats; and time spent on project activities. Post-meeting feedback surveys cover participant engagement, group dynamics, and usefulness of the intervention at each step. Pilot tests were conducted across combinations of device specifications to assess both functionality and usability. Functionality problems were fixed as they were documented. The average System Usability Scale scores for seven student testers were similar between the two interfaces: 72.9 for the mobile app and 72.5 for the website, both corresponding to "good" usability. Development of a mobile app requires substantial effort and personnel time which may not be apparent at the start of the project. Communication between researchers and the app developer was challenging at times. Conceptual and mental models of the specifications do not necessarily correspond due to differences in fields, experience, and priorities.

Keywords: Mobile applications · Usability testing · User-experience design

S. Bagnara et al. (Eds.): IEA 2018, AISC 824, pp. 523–530, 2019.
https://doi.org/10.1007/978-3-319-96071-5_56

1 Introduction

Process evaluation is valuable to assess quality, fidelity, and un-anticipated obstacles of a workplace intervention but can also be time-consuming and expensive. Numerous on-site visits and sources of information are often required to collect comprehensive process data; more efficient methods might enhance evaluation scope and completeness. Mobile applications ("apps") are an emerging data collection medium for research, especially for frequent real-time tracking. Tracking of ongoing intervention processes can identify factors that lead to intervention success or failure.

A mobile app was developed to collect process evaluation metrics for the "Safety and Health through Integrated, Facilitated Teams" (SHIFT) research study, an evaluation of the CPH-NEW Healthy Workplace Participatory Program in public sector healthcare institutions. Participants are members of joint labor-management health and safety committees, including committee co-facilitators, as well as internal program champions, other managers, and members of the research team.

Usability testing helps to ensure that the product meets the needs and capabilities of both the end-user and the researcher. The aims of this study are to describe a process to assess and (where necessary) improve functionality and usability of a process evaluation mobile app.

2 Methods

2.1 System Development

2.1.1 Mobile Application

An agile approach was utilized to develop a mobile application for both iOS (8.0+) and Android (4.1+) platforms. Detailed specs for the app were developed by the researchers, in consultation with a computer scientist. The app records four "functions:" meetings held, attendance and attendee feedback; project-related chats; and time spent on project activities (Fig. 1). Post-meeting feedback surveys (3 to 7 questions per meeting) cover participant engagement, group dynamics, and usefulness of the intervention at each step. Reminder push notifications for meetings are sent 2 and 24 h in advance. Post-meeting surveys are sent at the scheduled meeting end time, and reminders for incomplete surveys are sent after 2 and 24 h. A built-in background page-timer keeps track of app usage via Google Analytics (pending IRB approval).

2.1.2 Backend Website and Server

Co-facilitators may set meeting times and upload meeting minutes and agendas using the backend website. Incoming data are encrypted and sent to a password-protected server hosted at the University of Massachusetts Lowell, accessible to researchers via an accompanying website. Researchers are able to monitor data in real time and download data.

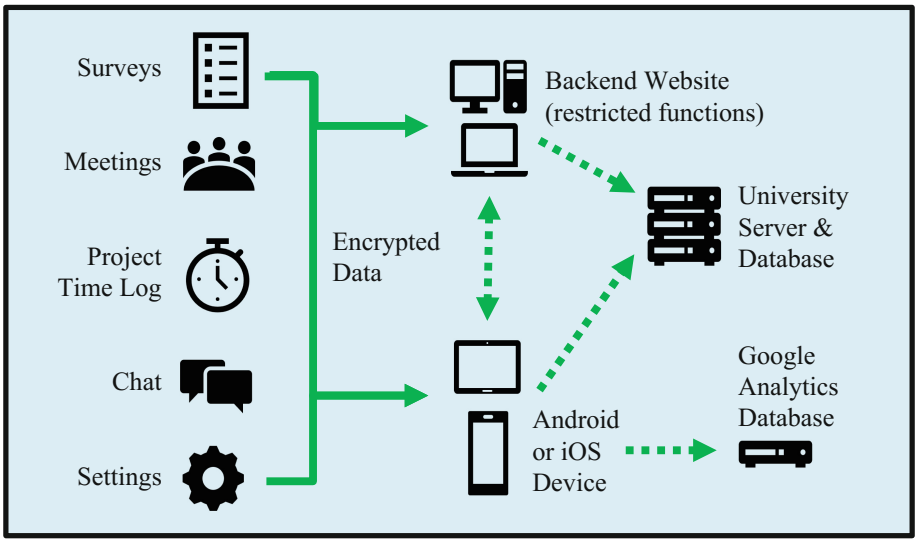

Fig. 1. System architecture of the SHIFT project mobile application

2.2 Usability Testing

Working prototypes of the app and website were developed and pilot-tested by university staff and students iteratively over calendar time as the app development progressed (Table 1). The first four testers evaluated the first prototype. One tester assessed the app when it was almost complete with all features, and the last two when it had full functionality.

Task scenarios were created to mimic the functionality expected from the app across roles, e.g., only committee co-facilitators and possibly researchers would set meeting times, whereas any employee at the healthcare facility might log project time. As the website will be used by only a few roles (co-facilitators and researchers), testers for the web tasks were assigned to those roles. Each individual carried out mock task scenarios and reported on ease of completing a scenario, whether errors were encountered, and whether a more effective method could be used. Usability for the app and website (if engaged during test) was rated separately using the System Usability Scale (SUS) [1]. Follow-ups were scheduled with testers to clarify feedback and issues. Pilot testing led to a much more refined prototype, which was then tested by selected end-users before full deployment.

Thematic analysis was carried out with student feedback and grouped as either functionality issues or usability concerns. The usability feedback was categorized by Nielsens' usability heuristics [2]. Where suggestions identified could not be implemented by the developer due to time-constraints or budget, alternative solutions proposed by the research team were documented and implemented.

Table 1. Test Devices Specifications

	Tester number						
	1	2	3	4	5	6	7
Test date	Dec. 2017	Jan. 2017	Jan. 2017	Jan. 2017	May 2017	Nov. 2017	Feb. 2018
App version	0.1.0	0.1.3	0.1.3	0.1.3	0.2.3	0.2.4-0.2.5	0.3.2
Phone platform	Android	iOS	Android	Android	iOS	iOS	Android
Phone manufacturer	Samsung	iPhone	Samsung	Samsung	iPhone	iPhone	Samsung
Phone model	Galaxy S7	6 s+	Galaxy Note 5	Galaxy S4 mini	5 s	5 s	Galaxy S7
Operating system version	6.0.1	10.2.0	6.0.1	4.4.2	8.1.3	11.1.2	7.0
Screen size	5.1”	5.5”	5.7”	4.3”	4”	4”	5.1”
Screen resolution	2560 × 1440	1080 × 1920	2560 × 1440	540 × 960	640 × 1136	640 × 1136	2560 × 1440
No. roles tested	7	4	4	4	1	7	7

3 Results

Pilot tests revealed functionality and usability concerns which were discussed between the research team and developer. Testing time ranged from 15 min to 5 h depending on whether the student had agreed to moderated or unmoderated testing and how many roles they were willing to test. Depending on the tester's availability, testing ranged from testing one to seven roles at a time, and may have been remotely tested or observed by the test facilitator. Observed student testers were requested to think aloud while performing task scenarios. A few student testers scheduled a second date in order to complete testing.

Most functionality issues were identified early after initial development and communicated quickly to the developer. All functionality issues were corrected before the app was completed. Some usability concerns (Table 2) were not rectified because of the time required.

3.1 Functionality Issues

Surveys: In tests of the first working prototype, almost half of the feedback for survey tasks stated that surveys did not work or partially worked. When student testers were asked whether they could complete survey task scenarios, some reported issues in viewing system status such as, *"there are no questions on the survey"* and *"shows surveys are loading and then blank."* One tester also reported, *"survey section of my app froze, so I could not see any sort of survey that might have come from the meeting."*

Other issues included receiving a survey in which *"I could only pick one multiple choice answer to answer for the surveys' three questions."* This would result in only one answer being reported while the other two would be missing.

To counter issues with loss of data connection, a feature was initially created to cache survey responses locally and send to the database when a connection was restored. Users reported issues where *"survey did not register as taken when I turned airplane mode off,"* which seemed to indicate that survey data would not be sent through.

Meetings: Most of the feedback stated that meeting creation could be completed, however a few problems occurred when creating a meeting in early prototypes, where *"nothing happened; I had to press cancel to get out of the new meeting window."* This lack of meeting creation could obviously result in loss of participation in site activities, communication issues, and loss of data. Additionally, certain *"members are not allowed to create meetings"* for meeting types that they should have been able to enter. A tester also reported *"can't cancel meetings."*

Regarding meeting reminders, one tester stated, *"I created a meeting one day and two minutes ahead of the present, after two minutes had passed, I did not receive any notification stating that I am 24 h away from a meeting. Any notification won't appear two hours before the meeting."*

Table 2. Usability feedback from student testers and research team response

Heuristic category	Example from feedback	Research team response
User control and freedom	*"I could not go back to the survey I completed and make edits to it."*	We judged it relatively unlikely that participants will frequently attempt to edit their responses, so we have decided to forgo this feature.
	"If you are in the middle of initial meeting-making, you can edit the date and time. However, as soon as you create the meeting, you cannot edit it, thus, in order to make the meeting correct, you need to 'cancel' the current meeting you have and create a whole new one with the correct date and time."	The many engineering-hours required to code this feature, coupled with an existing feature to cancel a mistake meeting, were reasons to forgo this feature. However, this was agreed to be something that would improve user experience and efficiency. While discussing this feature, the research team decided on an email notification feature to indicate when meetings were created and canceled, to assist with communicating changes outside the app, instead.
Consistency and standards	*"I'm not sure if the employees use a 24 clock but if they don't maybe change the times to a 12 h."*	Research team agreed with feedback. This was a relatively easy fix for the developer.
Recognition rather than recall	*"make it easier to figure out which survey is which"*	Research team agreed with feedback. Developer coded in a "subject" for meetings, which in turn would be the identifier for the associated survey.
Help and documentation	*"no agenda attached to it because the app does not give you that option."*	Task scenario was clarified so that this was not an expectation from the user.
	"security clearance to see time logs"	Task scenario was clarified so that this was not an expectation from the user for certain roles.

Project Time Log: Logging time spent on project activities is needed to track personnel cost of the intervention. Almost all feedback indicated success in logging project time. Some of the feedback stated that *"submitted project times don't show up under project time history"* and *"the time log history doesn't appear below the place where log in times, causing me to think that it never went through."* One tester then used the backend website and found that *"the time log history did appear on the web, but not on the app for I don't know what reason."* As many users will not be using the backend, the time-log history needs to be updated in real-time for users to know that their information was successfully submitted.

Chat: Most testers were able to chat successfully throughout all iterations of the app. There were certain times where *"no message received"* occurred for testers using the chat feature. Additionally, some testers reported *"unnotified outside the app and not receiving an immediately updated chat,"* indicating a problem with the notification feature.

3.2 Usability Issues

The main usability issues were inability to edit a created meeting and how to identify which meeting each survey belonged to. The examples from student testers (Table 2) provided insight on problems to fix before deployment to the end-user. While most issues led to re-design, a few would have been prohibitively expensive in personnel time.

3.3 System Usability Scale Scores

The average system usability scale scores for seven student testers were similar between the two interfaces: 72.9 for the mobile app and 72.5 for the website, both corresponding to "good" usability.

4 Discussion

Overall, early pilot tests revealed numerous functionality issues. These were most prominent in the survey and meeting functions, which was of utmost concern as these collect important process measures for the SHIFT project. As the meeting and survey functions were connected to each other, an issue with meeting creation could subsequently affect the survey function. These issues may also have been a result of changes in functionality requirements as the research project evolved, pinpointing functions that should be restricted to certain roles. New roles were also created. Limited phone use on site or smartphone ownership also led to requests for additional communication features.

Additionally, the layers within the research project – such as having different surveys for each type of role and having to sort respondent by site – were complex and resulted in the research team requesting more control over aspects of the software and data than the developer had initially understood. The developer spent substantial amounts of engineering time to restructure the backend and accommodate these changes, which in turn may have led to new functionality bugs.

Usability flaws did not seem to impact usability scores and may be attributed to the simple design of the app interface. Surprisingly, there were usability concerns discussed by the research team that were not mentioned by the student testers. This could be due to the limited scope of the test scenarios, which were purposefully limited, as on-site users are expected to have very limited time to devote to providing feedback to the app.

Pilot testing revealed that task scenarios were perhaps too vague, as a significant amount of time was spent after the usability test to reproduce bugs for the developer.

To remedy this, the task scenarios were revised for specificity and requested step-by-step information if a bug was encountered. This reduced the amount of follow-up, as on-site users will likely have less time to contribute to assisting with error investigation.

The development process might have been more efficient if the researchers had involved the developer at an earlier stage of the project, where specific features were discussed. Conceptual and mental models of the specifications did not necessarily correspond between the two parties due to differences in fields, experience, and priorities. In-person meetings and synchronous work may have also facilitated more efficient and effective communication than online methods.

Development and testing of a mobile app requires substantial effort which may not be apparent at the start of the project. Substantial time was needed to test across a diversity of mobile devices, platforms, networks, and software versions. Researchers looking to develop a mobile application for their research project should budget for personnel time needed to test for usability and software bugs. Buffer times should also be included in the timeline to avoid project delays.

Acknowledgements. This research was funded by Grant Number U19-OH008857, the National Institute of Occupational Safety and Health (NIOSH/CDC). The research team thanks the students and end-user volunteers who tested the mobile application and website.

References

1. Brooke J (1996) SUS: A “quick and dirty” usability scale. Usability Eval Ind 189–194:4–7
2. Nielsen J (1995) Usability heuristics for user interface design. https://www.nngroup.com/articles/ten-usability-heuristics/. Accessed 7 May 2018