Wheely

Usability Test Plan

v1

Anthony Driscoll
Monday, December 9, 2013

Table of Contents

Document Overview

This document describes a test plan for conducting a usability test during the development of Wheely. The goals of usability testing include establishing a baseline of user performance, establishing and validating user performance measures, and identifying potential design concerns to be addressed in order to improve the efficiency, productivity, and end-user satisfaction

The usability test objectives are:

- To determine design inconsistencies and usability problem areas within the user interface and content areas. Potential sources of error may include:
  - Navigation errors – failure to locate functions, excessive gestures to complete a function, failure to follow recommended screen flow.
  - Presentation errors – failure to locate and properly act upon desired information in screens, selection errors due to labeling ambiguities.
  - Control usage problems – improper toolbar or entry field usage.
- Exercise the application under controlled test conditions with representative users. Data will be used to access whether usability goals regarding an effective, efficient, and well-received user interface have been achieved.
- Establish baseline user performance and user-satisfaction levels of the user interface for future usability evaluations.

The targeted user groups include people with physical disabilities, elderly, and possibly families with children in strollers. The main target are people who are wheelchair bound and require the use of elevators in the subway system. There will be five participants, three in wheelchairs, one elderly, and one couple with children. The test will be performed outside subways stations and at disability activist group meet ups. The expected date range for this testing will be from 12/3/13 - 12/6/13.

Executive Summary

The usability of the map interface and task scenarios will be tested for Wheely. The participants will be given a task to find a specific station using the map, locate the available elevators for that station, and find directions to that elevator.

Upon review of this usability test plan, including the draft task scenarios and usability goals for Wheely, documented acceptance of the plan is expected.

Methodology

There will be five participants, evaluated outside a subway station or during a support group meet up. The use of two cameras will be used, one directed at the screen of the app to capture user interaction, and the other to be directed towards the face of the participant to record user response. Demographic and disability rating will be recorded. Notes will also be taken on the users satisfactory rating and any suggestions they may have for improvement.

Participants

*Due to lack of cooperating weather and some cancelled appointments, two able bodied participants have been included in the study.

Five participants are expected to participate and will be recruited outside subway stations or contacted directly to meet. Characteristics of their eligibility will include whether they are in a wheelchair, have trouble standing for long periods of time, or use strollers. They will also be asked whether or not the travel the subway system on a regular basis and if they use elevators at the stations. Participants will be expected to have some knowledge of applications and mobile devices. The age range will vary from 18-65 years old.

The participants' responsibilities will be to attempt to complete a set of representative task scenarios presented to them in as efficient and timely a manner as possible, and to provide feedback regarding the usability and acceptability of the user interface. The participants will be directed to provide honest opinions regarding the usability of the application, and to participate in post-session subjective questionnaires and debriefing.

The team will select test participants based on their subway and elevator usage. They will need to have knowledge of iOS applications and devices. Participants will need to be physically challenged in subway that forces them to use elevators. Preferably users in wheelchairs. Participants should also have some knowledge of the Metro Transit Authority subway maps.

Training

Participants will receive a general over overview of iPad gestures and will be shown the app in its open state due to prototype restrictions. The participants will receive and overview of the usability test procedure, equipment and software. The participants will be advised that this prototype is in early stages of development and many images and text are used as placeholders.

Procedure

Participants will take part in the usability test at various subway station entrances in New York City. An iPad with the application and supporting software will be used in a typical city environment. The participant may be asked to sit down with the moderator in a nearby square or park. The participant's interaction with the application will be monitored by the facilitator seated with the moderator. The test sessions will be videotaped.

The facilitator will brief the participants on the application and instruct the participant that they are evaluating the application, rather than the facilitator evaluating the participant. Participants will sign an informed consent that acknowledges: the participation is voluntary, that participation can cease at any time, and that the session will be videotaped but their privacy of identification will be safeguarded. The facilitator will ask the participant if they have any questions.

Participants will complete a pretest demographic and background information questionnaire. The facilitator will explain that the amount of time taken to complete the test task will be measured and that exploratory behavior outside the task flow should not occur until after task completion. At the start of each task, the participant will read aloud the task description from the printed copy and begin the task. Time-on-task measurement begins when the participant starts the task.

The facilitator will instruct the participant to 'think aloud' so that a verbal record exists of their interaction with the application. The facilitator will observe and video tape and record user behavior, user comments, and system actions using a dual camera approach.

After the given task, the participant will complete the post-task questionnaire and elaborate on the task session with the facilitator. After all task scenarios are attempted, the participant will complete the post-test satisfaction questionnaire.

Roles

The roles involved in a usability test are as follows. An individual may play multiple roles and tests may not require all roles.

Facilitator/Trainer

- Provide training overview prior to usability testing
- Provides overview of study to participants
- Defines usability and purpose of usability testing to participants
- Assists in conduct of participant and observer debriefing sessions
- Responds to participant's requests for assistance
- Records participant's actions and comments
- Provides overview of study to participants
- Defines usability and purpose of usability testing to participants
- Assists in conduct of participant and observer debriefing sessions
- Responds to participant's requests for assistance

Ethics

All persons involved with the usability test are required to adhere to the following ethical guidelines:

- The performance of any test participant must not be individually attributable. Individual participant's name should not be used in reference outside the testing session.
- A description of the participant's performance should not be reported to his or her manager.

Usability Tasks

The task descriptions below are required to be reviewed by the application owner, business-process owner, development owner, and/or deployment manager to ensure that the content, format, and presentation are representative of real use and substantially evaluate the total application. Their acceptance is to be documented prior to usability test.

The application's test setup will include a setting on the streets of NYC at a park or local square with seating as a more comfortable setting. The participant will be asked to find a specific station and its elevator location. Once the elevator location is found the moderator will ask the participant to find directions to that specific elevator. These tasks will support whether the functionality and interface of the map and information page is viable is useful.

A user scenario would include a wheelchair bound user with some type of physical disability forcing them to use elevators to access the subway system. This user may have trouble finding elevators or stations that offer elevator access due to busy streets and lack of city wide signage. The user may also have trouble using a normal subway map because of its fine print and detailed information.

The task being tested will display results of whether or not this application will be helpful in using the subway system for people in wheel chairs. The ability for the user to find elevators and accessible stations in the app will determine its usefulness.

Usability Metrics

Usability metrics refers to user performance measured against specific performance goals necessary to satisfy usability requirements. Scenario completion success rates, adherence to dialog scripts, error rates, and subjective evaluations will be used. Time-to-completion of scenarios will also be collected.

Scenario Completion

Each scenario will require, or request, that the participant obtains or inputs specific data that would be used in course of a typical task. The scenario is completed when the participant indicates the scenario's goal has been obtained (whether successfully or unsuccessfully) or the participant requests and

receives sufficient guidance as to warrant scoring the scenario as a critical error.

Critical Errors

Critical errors are deviations at completion from the targets of the scenario. Obtaining or otherwise reporting of the wrong data value due to participant workflow is a critical error. Participants may or may not be aware that the task goal is incorrect or incomplete.

Independent completion of the scenario is a universal goal; help obtained from the other usability test roles is cause to score the scenario a critical error. Critical errors can also be assigned when the participant initiates (or attempts to initiate) and action that will result in the goal state becoming unobtainable. In general, critical errors are unresolved errors during the process of completing the task or errors that produce an incorrect outcome.

Non-critical Errors

Non-critical errors are errors that are recovered from by the participant or, if not detected, do not result in processing problems or unexpected results. Although non-critical errors can be undetected by the participant, when they are detected they are generally frustrating to the participant.

These errors may be procedural, in which the participant does not complete a scenario in the most optimal means (e.g., excessive steps and keystrokes). These errors may also be errors of confusion (ex., initially selecting the wrong function, using a user-interface control incorrectly such as attempting to edit an un-editable field).

Noncritical errors can always be recovered from during the process of completing the scenario. Exploratory behavior, such as opening the wrong menu while searching for a function, will be coded as a non-critical error.

Subjective Evaluations

Subjective evaluations regarding ease of use and satisfaction will be collected via questionnaires, and during debriefing at the conclusion of the session. The questionnaires will utilize free-form responses and rating scales.

Scenario Completion Time (time on task)

The time to complete each scenario, not including subjective evaluation durations, will be recorded.

Usability Goals

This section describes the usability goals for Wheely.

Completion Rate

Completion rate is the percentage of test participants who successfully complete the task without critical errors. A critical error is defined as an error that results in an incorrect or incomplete outcome. In other words, the completion rate represents the percentage of participants who, when they are finished with the specified task, have an "output" that is correct. Note: If a

participant requires assistance in order to achieve a correct output then the task will be scored as a critical error and the overall completion rate for the task will be affected.

A completion rate of 100% is the goal for each task in this usability test.

Error–free rate

Error-free rate is the percentage of test participants who complete the task without any errors (critical **or** non-critical errors). A non-critical error is an error that would not have an impact on the final output of the task but would result in the task being completed less efficiently.

An error-free rate of 70% is the goal for each task in this usability test.

Time on Task (TOT)

The time to complete a scenario is referred to as "time on task". It is measured from the time the person begins the scenario to the time he/she signals completion.

Subjective Measures

Subjective opinions about specific tasks, time to perform each task, features, and functionality will be surveyed. At the end of the test, participants will rate their satisfaction with the overall system. Combined with the interview/ debriefing session, these data are used to assess attitudes of the participants.

Problem Severity

To prioritize recommendations, a method of problem severity classification will be used in the analysis of the data collected during evaluation activities. The approach treats problem severity as a combination of two factors - the impact of the problem and the frequency of users experiencing the problem during the evaluation.

Impact

Impact is the ranking of the consequences of the problem by defining the level of impact that the problem has on successful task completion. There are three levels of impact:

- High - prevents the user from completing the task (critical error)
- Moderate - causes user difficulty but the task can be completed (non-critical error)
- Low – minor problems that do not significantly affect the task completion (non–critical error)

Frequency

Frequency is the percentage of participants who experience the problem when working on a task.

- High: 60% or more of the participants experience the problem
- Moderate: 20% - 40% of participants experience the problem
- Low: 20% or fewer of the participants experience the problem

Problem Severity Classification

The identified severity for each problem implies a general reward for resolving it, and a general risk for not addressing it, in the current release.

Severity 1 - High impact problems that often prevent a user from correctly completing a task. They occur in varying frequency and are characteristic of calls to the Help Desk. Reward for resolution is typically exhibited in fewer Help Desk calls and reduced redevelopment costs.

Severity 2 - Moderate to high frequency problems with moderate to low impact are typical of erroneous actions that the participant recognizes needs to be undone. Reward for resolution is typically exhibited in reduced time on task and decreased training costs.

Severity 3 - Either moderate problems with low frequency or low problems with moderate frequency; these are minor annoyance problems faced by a number of participants. Reward for resolution is typically exhibited in reduced time on task and increased data integrity.

Severity 4 - Low impact problems faced by few participants; there is low risk to not resolving these problems. Reward for resolution is typically exhibited in increased user satisfaction.

Reporting Results

The Usability Test Report will be provided at the conclusion of the usability test. It will consist of a report and/or a presentation of the results; evaluate the usability metrics against the pre-approved goals, subjective evaluations, and specific usability problems and recommendations for resolution. The recommendations will be categorically sized by development to aid in implementation strategy. The report is anticipated to be delivered to the Project UCD Contact by December 9, 2013.