# Green Envy Application Usability Test

*Sami Al-Qusus*
*CSCI310*
*VIU, Nanaimo, Canada, V9R 5S5*
*Email: sami.qusus@gmail.com*

*December 5, 2017*

# Abstract

The aim of this report is to formally communicate and report the outcome of a usability test, its results and thereof analysis. The usability test ought to validate Green Envy's interface design for its initial "ease of use" proposition. The team conducted an in lab usability test following the think-aloud protocol, by gathering quantitative and qualitative data collected from four subjects. In general all participants thought that the design was easy to use and follow, while quantitative results have shown slightly less conformance. Concluding that three components need to be redesigned accordingly.

# Table of Contents

# 1. Introduction

When a product is designed for public service, it's important that ease of use, usability requirement, is given priority in such an application. Since, Green Envy's software is meant to provide recycling information in a simple interactive way and is to be integrated with Nanaimo's Garbage and Recycling application, that's intended for public service, its important that the application is tested for how well it complies with the ease of use usability requirement.

Therefore, usability testing was conducted using Green Envy's latest prototype, to validate the extent of Green Envy's said "ease of use" proposition. The team conducted an in lab usability test following the think-aloud protocol, by gathering quantitative and qualitative data collected from four participants. A laptop captured audio and screen recordings of each participant's speech and navigation patterns. Participants then filled out a follow up questionnaire at the end of the session, for qualitative data collection purposes.

# 2. Methodology

The team conducted an in lab usability test following the think-aloud protocol, by gathering quantitative and qualitative data collected from four participants, checking for time taken and number of errors that result on each test.

## 2.1 Procedure

One of each of the four-team members was able to recruit one participant for the study. Each recruiting member contacted one participant personally. Upon receiving a signed consent form from recruits, team members coordinated with their recruit setting a time, date, and place to conduct the test. Before meeting with the participant, the tester had all materials required for the test prepared to avoid inconvenience to participant. Each test session lasted around 15 minutes, including the test and collecting all required documents and feedback from participant. During the session, the tester first requested the participant to fill out a demographics questionnaire (attached in Appendix I), and then explained to the participant the tasks requested of them and instructed them to attempt one test at a time. Each of the four subjects performed the tasks asked of them, performing tasks in the order they wish, randomizing the order to try and limit the effect of a learning curve. Finally, once participants went through all the tasks, they were asked to fill out a post-test questionnaire (attached in Appendix I) and for any constructive criticism or helpful comments.

## 2.2 Participants

Although at least five participants are usually needed in order to obtain dependable results, only four were tested, because the experiment was relatively small. All four subjects needed to be relatively competent with smartphone use, because the application developed is targeting moderate smartphone users. Demographics of relevance for the four subjects are presented bellow.

| ID | s1 | s2 | s3 | s4 |
|---|---|---|---|---|
| Age | 54 | 22 | 60 | 24 |
| Occupation | Head chef | VIU Student | Mother | VIU Student |
| Smartphone Competence | Competent | Competent | Competent | Competent |

## 2.3 Setting & Materials

For setting, participants completed the test in an in-lab environment, either in the labs at VIU or in a similar environment at home or at a public place depending on their convenience as long as the setting was a quiet and comfortable one.

The materials used for conducting the usability test consisted of a Laptop with audio and screen recording capabilities to record speech and capture participant's mouse navigation throughout the test, a PDF prototype of Green Envy's application for participants to apply tasks requested of them, Microsoft Excel to analyze quantitative data collected, four consent forms for participants to understand the study's purpose and procedure and having signed consent, four questionnaires for qualitative data and collecting participant's feedback, four questionnaires to collect demographics, and lastly four stapled copies of instructions and tasks for users to follow while conducting the test.

The test material consisted of test items required for both the quantitative and qualitative parts of the test and material required for later analyzing of such data.

Attachments of questionnaires are included in Appendix I.

## 2.4 Tasks

The tasks the participants performed are shown bellow.

**Directions on how to run the prototype.**

**Scenario 1 Overview of Steps:**
1. Open Green Envy app
2. Go to BLOGS
3. Search for "tire repurposing"
4. if no results are found create a new blog
   .      a)  give the blog the title "tire repurposing"
   .      b)  Add "How do you repurpose tires?" in blog's body
5. choose notify me for comments
6. Post the blog
7. Check notifications.
   .      a)  read new comment
8. close app

**Scenario 2 Overview of Steps:**
1. Open Green Envy app
2. Search if cardboard is recyclable
   .      a)  search "cardboard"
   .      b)  go to "cardboard"
   .      c) read if cardboard is recyclable or not
3. check for "depot locations near you"
4. close app

**Scenario 3 Overview of Steps:**
1. Open Green Envy app
2. Go to GAME
3. add "4" items to collect points for.
4. close app

## 2.5 Measures

Quantitative data was measured by collecting the time each participant took to complete one task and by colleting the error count that was measured by reviewing the screen recordings for each task and counting how many times a user attempted to click on the wrong button or field. Qualitative data was collected by requesting each participant to fill out a post-test questionnaire (attached in Appendix I) and for any constructive criticism or helpful comments participants provided at the end of the test session. Qualitative data was also collected by reviewing all user's think-aloud task recordings for any unaccounted for interpretation of the tasks.

## 2.6 Problems

One of the testers forgot to tell a participant to speak out loud while performing the task and had to get the user to restart the test. This problem should be avoided in future iterations as much as possible, because it might lead to a learning curve advantage in one of the tests without it being accounted for.

# 3. Results

## 3.1 Quantitative Results

Bellow is the qualitative data collected. Provided are the times each participant took to complete one task and the error count that was measured by reviewing the screen recordings for each task and counting how many times a user attempted to click on the wrong button or field.

| Subject | Task | Errors | Time (seconds) |
|---|---|---|---|
| s1 | 1 | 2 | 83.361 |
| s2 | 1 | 1 | 102.568 |
| s3 | 1 | 0 | 98.217 |
| s4 | 1 | 3 | 87.753 |
| | | | |
| s1 | 2 | 1 | 84.342 |
| s2 | 2 | 1 | 93.633 |
| s3 | 2 | 2 | 92.772 |
| s4 | 2 | 0 | 85.896 |
| | | | |
| s1 | 3 | 2 | 90.364 |
| s2 | 3 | 1 | 100.382 |
| s3 | 3 | 0 | 96.775 |
| s4 | 3 | 2 | 92.565 |

| Average of All 3 Tasks | |
|---|---|
| s1 | 86.022 |
| s2 | 98.861 |
| s3 | 95.921 |
| s4 | 88.738 |

Mean for Task 1 is, 92.97
Mean for Task 2 is, 89.16
Mean for Task 3 is, 95.02
Grand mean = 92.39, the overall mean time taken.

Mean error for all Tasks done by s1, 1.67
Mean error for all Tasks done by s2, 1
Mean error for all Tasks done by s3, 0.67
Mean error for all Tasks done by s4, 1.67

Mean error for Task 1 is, 1.5
Mean error for Task 2 is, 1
Mean error for Task 3, is 1.25
Grand mean = 1.25, the overall mean of errors.

## 3.2 Qualitative Results

Bellow is the overall qualitative data outcome that was collected from post-test questionnaires, constructive criticism or helpful comments that participants provided at the end of the test sessions and the qualitative data that was collected by reviewing all user's think-aloud task recordings and noting all unaccounted for interpretations of tasks.

Post-Test Questionnaire:

All users consider themselves competent with smartphones. All users stated that they would use the application if they needed to, but two said that they are pretty familiar with the recycling process and probably wouldn't ever need to use the application. They all said, they felt the application was relatively easy to use.

Constructive criticism/feedback/comments:

None of the users felt like there is anything further to report.

Reviewing all user's think-aloud task recordings and noting all unaccounted for interpretations of tasks:

Users across the board seemed to be hesitant when they are supposed to check the "notify me for comments" before they post a blog. Screen recordings, show that the mouse alternates between the "post" button and the "notify me for comments" selector, before participants finally select "notify me for comments". Participants have been found to stop talking and the mouse is observed circling the screen after participants state that they are looking for the notification alert. Final unaccounted for observation, is that participants on average ignore the comment that they should locate after clicking on notifications.

# 4. Discussion

## 4.1 Interpretation

For total time taken with respect to different subjects, we cannot see any correlation between any of the means and subjects demographics. Therefore, it would be fair to take, 92.39s, the grand mean time taken as an average time it would take future users on average to complete all tasks. As for total the time taken to accomplish one task compared to other tasks, we can see that all tasks are all relatively within the grand mean. Therefore there isn't anything that stands out, which might invalidate the ease of use proposition with regards to time.

As for error analysis, we can see that there is also no clear correlation between error and the age of participants. Two participants from two different age groups had a, 1.67, total mean error across the completed tasks, which was above the grand mean for error, 1.25. While the other two participants all from different age groups had total means bellow that of the average. As for total the error taken to accomplish one task compared to other tasks, we can see that all tasks are all relatively within the grand mean. Therefore there isn't anything that stands out, which might invalidate the ease of use proposition with regards to error. Concluding, that quantitatively the ease of use proposition claimed by the application holds.

As for qualitative findings, the three observations described in regards to the "notify me for comments", notification alert, and not clear comment feedback hinder certain aspects of the application unless appropriate fixes are put in place.

## 4.2 Fix

One possible fix that conforms with standards for the observed "notify me for comments" usability problem, would be to have the "notify me for comments" automatically selected, and unselected by the user if they require otherwise. Another fix that better follows usability standards would be to replace the exclamation mark icon in the notification alert with the digit that represents the notification count.  Finally, to better map a notification to a new comment while providing proper feedback is to focus on the new comment for a few seconds enough for the users eye to recognize and locate feedback.

## 4.3 Critical reflection

Findings might not be very accurate, because only four subjects were tested and the software was only an interactive PDF prototype.

## 4.4 Research agenda

If possible in the future, this research should be revisited and tested on a bigger number of subjects after the advised fixes have been considered.

# 5. Conclusion

Although results indicate that there isn't anything that stands out, which might invalidate the ease of use proposition with regards to time and error counted qualitatively, and hence quantitatively the ease of use proposition claimed by the application holds, qualitative findings show otherwise. Qualitative findings suggest that certain aspects of the application, unless appropriate fixes are put in place, do not completely comply with the ease of use usability requirement.

# References

Greenberg, Saul. *How to Structure Reports on Experiments in Human- Computer Interaction*. University of Calgary.

Trulock, V. "Understanding HCI." *Performance Measurement*, hci.ilikecake.ie/eval_performance.htm.

# Appendix I

Pre-Test Questionnaire
1. Age?
2. Occupation?
3. How competent are you with smartphones?

Post-Test Questionnaire
1. In general, do you think an app like this would be used? Yes/No
2. On a scale of 1 to 10. The overall look of this mobile app.
3. On a scale of 1 to 10. The ease of use of this mobile app.
4. Were there any difficulties using this app? Yes/No. If so what was it?
5. Is there anything that the app is missing, in your opinion?