My Med Schedule: Comparison Study

Table of Contents

- Executive Summary
  - Significant Findings
    - Time
    - Error Count
    - Clicks
- Purpose and Test Objective
  - Purpose
  - Test Objective
    - Hypothesis 1
    - Null Hypothesis 1
    - Hypothesis 2
    - Null Hypothesis 2
    - Hypothesis 3
    - Null Hypothesis 3
- Methodology
  - What Happened During the Comparison Test
  - Who We Tested
  - What We Prepared for the Comparison Test
    - Perfect Path for Design 1 (Condition 1)
    - Perfect Path for Design 2 (Condition 2)
  - What Participants Did
  - What Data We Collected
  - Where We Tested
  - Test Environment
- Results
  - Time
  - Error Count
  - Clicks
- Findings & Recommendations
  - Count of Time Spent Online Daily
  - Success / Failures
  - Lessons Learned
  - Recommendations
- Principal Roles
- Appendix A: Collected Data
- Appendix B: Consent Form Template
- Appendix C: Testing Script
- Appendix D: T-Test Output Snapshots

My Med Schedule: Comparison Study

Executive Summary

The purpose of this comparison study is to test if changing the button layout of a perceived confusing web page will result in a better user experience and allow for a faster completion of the task. Using a between-groups study, the data we gathered supported our hypotheses that participants completed the activity in the new design faster than the original design, had fewer error counts, and performed fewer clicks to complete the activity.

Significant Findings

From 30 Participants tested, we gathered the following statistically significant findings between the two designs.

Time

Participants using the revised web page completed the activity faster (M=64, SD=55.2) than those who used the original design, (M=123, SD=82.3), t(2.31)=28, p=0.029.

Error Count

Participants using the revised web page had fewer error counts (M=0.8, SD=1.15) than those who used the original design, (M=4, SD=4.12), t(2.9)=28, p=0.007.

Clicks

Participants using the revised web page performed fewer clicks (M=3.13, SD=1.55) than those who used the original design, (M=6, SD=3.87), t(2.66)=28, p=0.013.

Purpose and Test Objective

Purpose

During our review of the MyMedSchedule.com application, our group observed participants becoming lost when performing the second task of scheduling medications reminders. One button that stood out was “Refill Reminders.” Participants inadvertently clicked the button and when they did, they lost focus on what to do next. It is our goal in this research project to understand whether a simple change of location of the buttons could eliminate or lessen the confusion and help individuals complete the task faster.

Test Objective

The objective of our comparison test will be to capture time taken, number of errors, and number of clicks for both the original refill reminder process and a revised process. These values will allow us to evaluate

the designs and identify if the modified design will be worth suggesting as a change for the application. Our initial expectations are the participants will select the wrong refill reminder button in the existing design. We also expect the new suggested web page will facilitate faster completion of task and prevent participants from becoming lost.

Hypothesis 1
Participants will complete the activity in the new design faster than the original design.

Null Hypothesis 1
Participants will not complete the activity in the new design faster than the original design.

Hypothesis 2
Participants will complete the activity with less error count in the new design than the original design.

Null Hypothesis 2
Participants will not complete the activity with less error count in the new design than the original design.

Hypothesis 3
Participants will complete the activity with less clicks in the new design than the original design.

Null Hypothesis 3
Participants will not complete the activity with less clicks in the new design than the original design.

Methodology

What Happened During the Comparison Test

Our group conducted a between-groups comparison test with the MyMedSchedule.com refill reminder process from November 4 to 7, 2015. Two members tested eight participants each and a third member tested fourteen participants for a total of thirty participants. Fifteen participants completed the task while using Mockup 1 while another fifteen participants used Mockup 2 during their task.

Before the testing, our group requested the consent of the participants through the use of a consent form (Appendix B) and verbal consent to video record the session. The participants who matched the profile were then asked to spend less than 10 minutes with one of the two prepared digital mockups. During this time, participants:

- Answered a question regarding online usage.
- Performed a refill reminder task in one of the mockups.

My Med Schedule: Comparison Study

Who We Tested

Thirty participants evaluated the MyMedSchedule.com refill reminder process. The participants had the following profile characteristics:

- Age ranged from 20 to 75 years old.
- Were familiar with completing a web form.

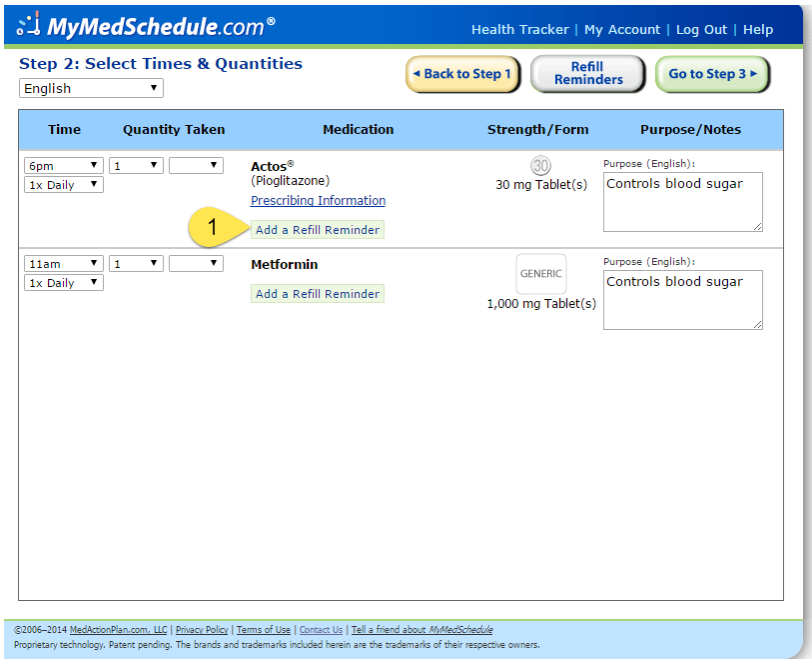
What We Prepared for the Comparison Test

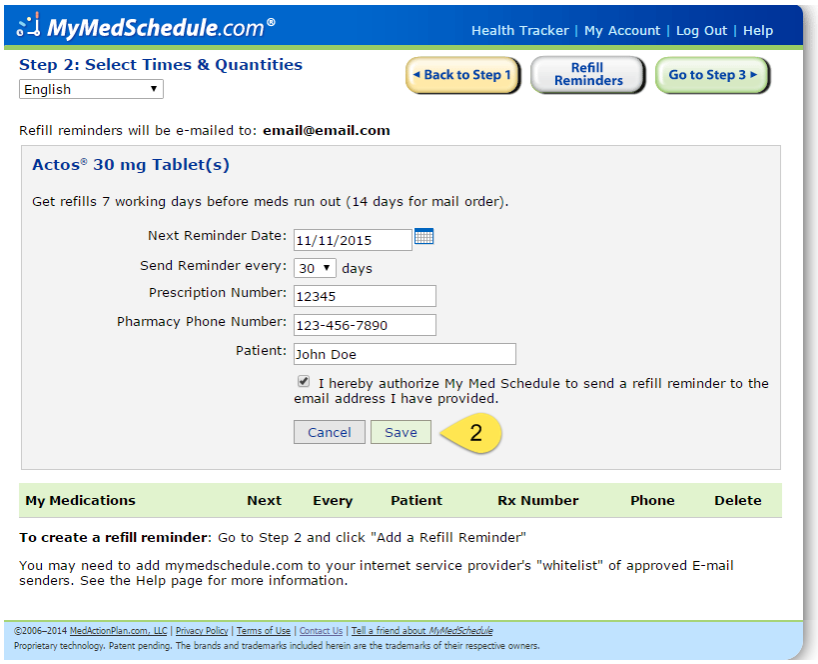
Two digital mockups were created for comparison.

- Condition 1: Existing MyMedSchedule pages
- Condition 2: Redesigned MyMedSchedule pages

The first mockup was created using the existing Step 2 screen layout from MyMedSchedule.com. The mockup had clickable links to other pages related to the task and any immediate navigation action, such as Step 1 or Step 3. These clickable links - not a part of the perfect path - were included to help simulate what would occur if participants decided to navigate around the immediate area of the site.

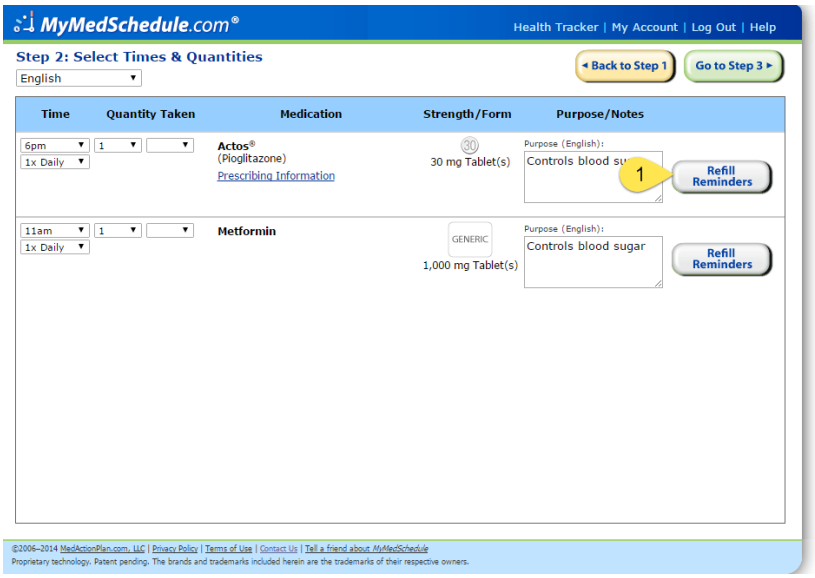
Perfect Path for Design 1 (Condition 1)

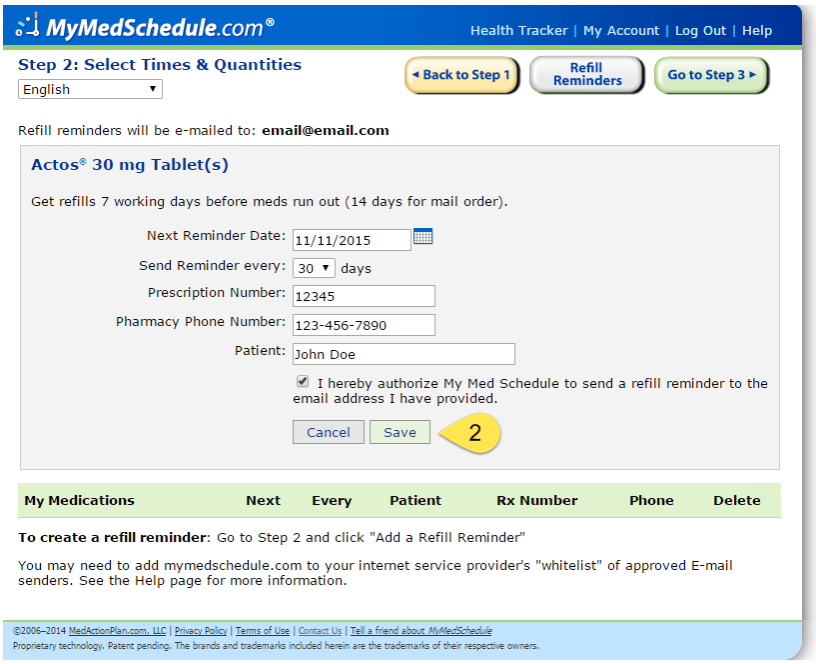
Click 1: "Add a refill reminder" Button	

Click 2: “Save” Button	

The secondary mockup was a modified version of Step 2 from MyMedSchedule.com. The top “Refill Reminder” button was removed and used as a template for clickable refill reminder actions and placed to the right of the medications in the list.

Perfect Path for Design 2 (Condition 2)

Click 1: “Refill Reminders” Button	

Click 2: “Save” Button	

In both mockups, data were prefilled in the forms to help participants focus on the main actions of the task. None of the form fields were set up to be clickable in the digital mockup.

What Participants Did

The participants were setup with one out of two digital mockups and asked to perform a single task. The first scenario was to use MyMedSchedule.com scheduler to do the following:

- Task: Create a medication refill reminder for Actos.

They were then asked their thoughts on activity. Please see the testing scripts in Appendix C to further review the process.

What Data We Collected

Focusing on our research question, the group captured these quantitative metrics data:

- Time Spent on Task
- Clicks
- Error Counts
- Success, Difficulty, or Failure condition

My Med Schedule: Comparison Study

For this report, we did not include qualitative responses from the participants after they completed the task. Our group wanted this report to focus on the observations of the task as it related to the quantitative data.

Where We Tested

Following is a summary of the computing environment participants used for the testing:

URL of tested website:	http://leo.primero.ws/hci460/MyMedSchedule
Browser tested:	Internet Explorer
Screen resolution:	At least [1024 X 768]
Operating system:	Windows 8/10
Connection speed:	Fast WiFi

Test Environment

For in-person testing, the test was conducted in a small conference room or private room that was available. A laptop with a mouse and keyboard were provided for completing the task. A wireless high speed wifi was used to connect to the digital mockups on the internet. In the case of no internet connection, a local copy of the digital mockups were used. A PC resolution of at least 1024x768 and the browser Internet Explorer 8/10.0 were used. Camtasia was used as the recording software to capture mouse movements and clicks, keyboard clicks, and any conversation that took place between the moderator and the participant.

For remote participants, the participants performed the test from their home. A video conference call (GotoMeeting) was used to record the interaction between moderator and the participants, including mouse and keyboard movements.

In both forms of testing, moderators acquired verbal permission before recording the participants activities and comments.

Results

Our group used an unpaired t-Test to analyze our between-groups data. We used an online tool to run the calculations (http://www.physics.csbsju.edu/stats/t-test_NROW_form.html).

While reviewing our data, we note Condition 1 was the test with the existing refill reminder layout design. Condition 2 was the test with the revised refill reminder layout.

My Med Schedule: Comparison Study

Time

Our group defines the total time to complete the task based on these starting and ending actions.

- Start time: When participants acknowledged the activity start, began looking at the screen, and moving the mouse.
- End time: When participants clicked the Save button to save the refill reminder.

Grade Level	n	Mean	SD	t	df	p
Condition 1	15	123	82.3	2.31	28	0.029
Condition 2	15	64	55.2			

Table 1: t-Test results comparing Condition 1 (existing design) and Condition 2 (revised design) time completion of activity.

Among participants who completed the activity (n = 30), there was a statically significant different between the two designs, Condition 1 (M = 123, SD = 82.3) and Condition 2 (M = 64, SD = 55.2), $t(28) = 2.31$, $p \leq .05$. Therefore, we were able to reject the null hypothesis that participants will not complete the activity in the new design faster than the original design.

Error Count

Our group defines error count as the clicks that were irrelative to complete the activity or took the participant away from the perfect path.

Grade Level	n	Mean	SD	t	df	p
Condition 1	15	4	4.12	2.9	28	0.007
Condition 2	15	0.8	1.15			

Table 2: t-Test results comparing Condition 1 (existing design) and Condition 2 (revised design) error count of activity.

Among participants who completed the activity (n = 30), there was a statically significant different between the two designs, Condition 1 (M = 4, SD = 4.12) and Condition 2 (M = 0.8, SD = 1.15), $t(28) = 2.9$, $p \leq .05$. Therefore, we were able to reject the null hypothesis that participants will not complete the activity with less error count in the new design than the original design.

Clicks

Our groups defines clicks as the mouse clicks participants made during the course of activity.

Grade Level	n	Mean	SD	t	df	p
Condition 1	15	6	3.87	2.66	28	0.013
Condition 2	15	3.13	1.55			

Table 3: t-Test results comparing Condition 1 (existing design) and Condition 2 (revised design) on activity clicks.

Among participants who completed the activity (n = 30), there was a statically significant different between the two designs, Condition 1 (M = 6, SD = 3.87) and Condition 2 (M = 3.13, SD = 1.55), $t(28) = 2.66$, $p \leq .05$. Therefore, we were able to reject the null hypothesis that participants will not complete the activity with less clicks in the new design than the original design.

Findings & Recommendations

Based on our results data and the statistical significance from the t-Tests for our three hypotheses, we propose there is a case for revising the layout of Step 2. Our revised layout shortened the length of time to complete the task and the number of errors participants made while attempting the task. We believe the improvement was due to these factors:

- Removing buttons with different actions but sharing the same “Refill Reminder” text reference.
- Modifying the button style by increasing the size of the text and highlight color around the button.
- Moving the button such that it was located in the same table row of as the medication, but was contained in its own column, not sharing a column with the medication name.

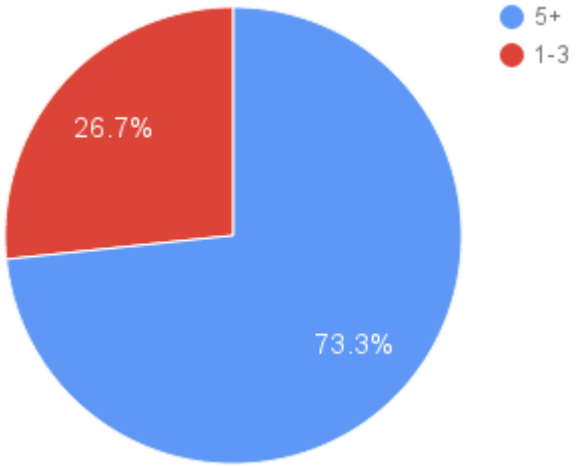
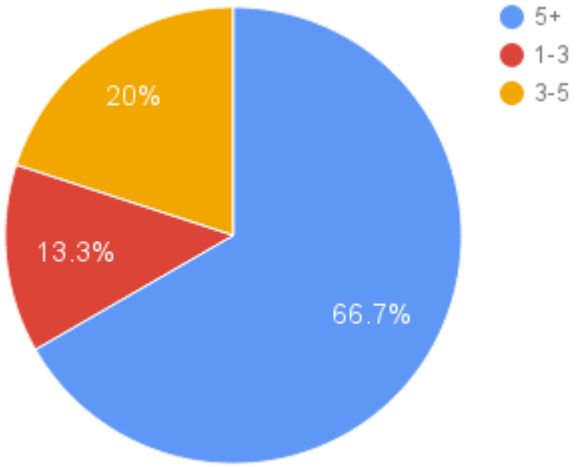
This potentially suggests the issues with the existing design are related to the visibility of the button and confusion of having multiple buttons on the screen contain similar text.

Here are some additional findings from our testing.

Count of Time Spent Online Daily

Our group found there was little difference between the two participant groups in terms of daily online use.

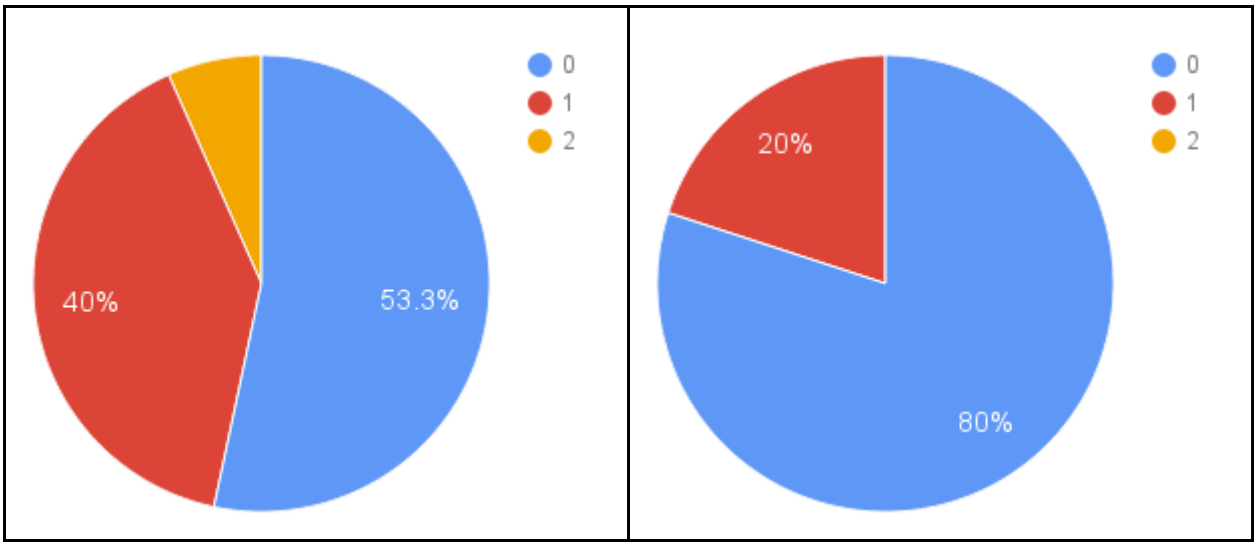
My Med Schedule: Comparison Study

CONDITION 1	CONDITION 2
This group has 4 participants (26.7%) who spend 1-3 hours online daily and 11 participants (73.3%) spending 5+ hours online daily.	This group has 2 participants (13.3%) who spend 1-3 hours online daily, 3 participants (20%) who spend 3-5 hours online daily and 10 participants (66.7%) spending 5+ hours online daily.
	

Success / Failures

The data our group collected on success and failure of the task aligned with our expectations as well as with the error count and clicks results data. The participant group for Condition 1 experienced more difficulty completing the task compared to the participant group for Condition 2.

CONDITION 1	CONDITION 2
This group has 1 participant (6.7%) who failed the task, 6 participants (40%) who needed significant help to finish the task, and with 8 participants (66.7%) successfully finishing the task.	This group completed the task as required with 3 participants (20%) who needed significant help to finish the task, and with 12 participants (80%) successfully finishing the task.



Lessons Learned

Our group encountered challenges when trying not to prompt participants and lead them to the correct path while performing the test. As the test went on, some participants became anxious and made extra error clicks or asked if they were correctly completing the path. For the most part, we were able to help the participants refocus by asking them to review or reread the scenario instructions and application labels.

There are a few situations which came up during our testing process that could have affected our results. One part was our limited screening process; we did not screen our participants for their level of technology proficiency (novices vs experts) which led us to testing participants with stronger computer skills, a few of whom were medical domain experts. Another part that affected participants actions during the testing was our disabling the data entry elements and providing default values. This caused some participants to express their confusion and ask why they could not change the values, slowing a portion of participants down. The last aspect that could have affected our testing was the starting point of the task. Some participants mentioned feeling lost because they were started at Step 2 of the full scheduling process, not at Step 1, the standard starting point when first logging into the application.

In light of these situations, there are a few actions our group could take to improve. Our group could benefit from additional practice with moderating tests. This should help strengthen our skills and prevent possible leading of participants in the future. We also need to spend additional time to define better screening rules, taking into account whether a person is or is not a domain expert or if the person is a novice or expert with computers. We acknowledge that having both types of expertise levels are okay to have, but we should make sure to track expertise level as a data point to help with balancing and

interpreting the data. This would help ensure we have a better balanced representation of different skillsets.

Recommendations

We would like to make the recommendation that the layout for Step 2 be modified to match or be similar to the revised design we created for our test. Making the refill reminder button larger and have its own column appeared to improve task completion time and reduce errors. If the design were to change to this layout or a close facsimile, then we believe the user experience would improve.

We have observed strong indications that the new design we devised will help users of MyMedSchedule.com have a more efficient and less confusing refill reminder process; however testing this process is just the start. Our group recommends additional testing to further validate these findings. In addition, there are still more sections of the web site which require research and testing to improve the functionality for their users. It is our hope that additional testing and research will be pursued to lessen users' frustrations with respect to unfavorable button placement.

Principal Roles

The following is a breakdown of the principal roles of the group in completing Project 4.

Leo Primero (leo@primero.ws)

- Created the application mock-ups
- Prepared the Executive Summary sections of the report
- Also worked on the Findings and Recommendation sections of the report

Hanna Ho (greenmice3@gmail.com)

- Devised the test scripts
- Worked on the Result section of the report
- Prepared the informal video presentation

Jason Boggs (jboggs.box@gmail.com)

- Developed the consent form for participants
- Prepared the Purpose, Test Objective, and Methodology sections of the report
- Ran the T-Testing process on the collected data

In addition to these primary roles, each member moderated the test with at least eight participants and contributed to the content of the report.

Appendix A: Collected Data

		Condition 1 - Current Interface					
Participant	P# Mod	Success(0) / Difficulty (1) / Failure(2)	Total Task (mins)	Total Task (secs)	Error Count	Clicks	Online Time (days)
1	P1H	0	0.47	28	0	2	5+
2	P2H	1	1.65	99	13	15	5+
3	P3H	0	0.75	45	10	12	5+
4	P4H	0	0.75	45	2	4	5+
5	P1L	1	3.29	197	7	6	5+
6	P3L	2	2.67	160	5	8	5+
7	P5L	1	2.19	131	1	4	5+
8	P7L	0	1.01	61	0	2	1-3
9	P10L	0	2.05	123	2	4	5+
10	P12L	1	5.70	342	7	9	1-3
11	P14L	1	3.72	223	8	6	1-3
12	P17J	1	1.77	106	3	8	5+
13	P18J	0	1.15	69	0	2	5+
14	P19J	0	2.14	128	0	2	1-3
15	P20J	0	1.47	88	2	6	5+
			Total	**1845**	**60**	**90**	

		Condition 2 - Suggested Interface					
Participant	P# Mod	Success(0) / Difficulty (1) / Failure(2)	Total Task (mins)	Total Task (secs)	Error Count	Clicks	Online Time (days)
16	P5H	0	0.35	21	2	4	5+
17	P6H	0	0.63	38	3	5	5+
18	P7H	1	1.55	93	1	3	5+
19	P8H	0	0.47	28	0	2	5+
20	P2L	0	1.70	102	0	3	5+
21	P4L	1	3.32	199	2	5	1-3
22	P6L	0	0.88	53	0	2	3-5
23	P8L	0	0.34	20	0	2	1-3
24	P9L	0	0.35	21	0	2	5+
25	P11L	0	0.74	44	0	2	5+
26	P13L	0	0.90	54	0	2	3-5
27	P21J	0	0.35	21	0	2	5+
28	P22J	1	2.78	167	3	7	5+
29	P23J	0	0.45	27	0	2	3-5
30	P24J	0	1.20	72	1	4	5+
			Total	**960**	**12**	**47**	

Appendix B: Consent Form Template

ADULT CONSENT TO PARTICIPATE IN RESEARCH

Comparison Testing of MyMedSchedule.com

Principal Investigator: [insert_name_here], MS candidate in Human Computer Interaction at DePaul University.

Institution: DePaul University, Chicago, Illinois, USA

College: College of Computing and Digital Media

Faculty Advisor: Enid Montague, PhD; DePaul University College of Computing and Digital Media

Collaborators: Leo Primero III, Hanna Ho, Jason Boggs

What is the purpose of this research?
We are asking you to be in a research study because we are trying to compare two versions of the MyMedSchedule.com Refill Reminder process. This study is being conducted by [insert_name_here], a graduate student at DePaul University as a requirement to obtain [his_or_her] Master's degree. This research is being supervised by [his_or_her] faculty advisor, Enid Montague. There may be other people on the research team assisting with the study. We hope to include about twenty four people in the research.

Why are you being asked to be in the research?
You are invited to participate in this study because you are an individual who is:

- Between the ages of 20 to 75.
- Are familiar with completing a web form

What is involved in being in the research study?
If you agree to be in this study, being in the research involves completing the refill reminder process through use of a digital mock-up of the MyMedSchedule.com web application.

The activity will be video recorded and reviewed later in order to get an accurate record of the actions you took while you used the MyMedSchedule.com mock-up.

How much time will this take?
The entire study will take less than 10 minutes to complete. The testing will involve an introduction of the task, completion of the task, and a short debriefing.

Are there any risks involved in participating in this study?
Being in this study does not involve any risks other than what you would encounter in daily life. Personal information recorded will be abstracted so that no identifying information will be published in any report. (See below for information regarding confidentiality.)

Are there any benefits to participating in this study?

You will not personally benefit from being in this study. We hope that what we learn will help improve the the functionality of the MyMedSchedule.com application for current and future users of the application.

Can you decide not to participate?
Your participation is voluntary, which means you can choose not to participate. There will be no negative consequences, penalties, or loss of benefits if you decide not to participate or change your mind later and withdraw from the research after you begin participating.

Who will see my study information and how will the confidentiality of the information collected for the research be protected?
The research records will be kept and stored securely. Your information will be combined with information from other people taking part in the study. When we write about the study or publish a paper to share the research with other researchers, we will write about the combined information we have gathered. We will not include your name or any information that will directly identify you. We will make every effort to prevent anyone who is not on the research team from knowing that you gave us information, or what that information is. However, some people might review or copy our records that may identify you in order to make sure we are following the required rules, laws, and regulations.

The video recordings will be kept until accurate written notes have been made, then they will be destroyed.

You should know that there are some circumstances in which we may have to show your information to other people. For example, the law may require us to show your information to a court or to tell authorities if you report information about a child being abused or neglected or if you pose a danger to yourself or someone else.

Who should be contacted for more information about the research?
Before you decide whether to accept this invitation to take part in the study, please ask any questions that might come to mind now. Later, if you have questions, suggestions, concerns, or complaints about the study or you want to get additional information or provide input about this research, you can contact the researcher:

Name: **[insert_name_here]** Phone: **[insert_phone_number]** Email: **[insert_email_here]**

You will be given a copy of this information to keep for your records.

Statement of Consent from the Subject:

I have read the above information. I have had all my questions and concerns answered. By signing below, I indicate my consent to be in the research.

Signature:_____

Printed name: _____

Date: _____

Appendix C: Testing Script

Test Scripts

Introduction (for both remote and live usability tests)

Note to moderator: Display a clean page (an empty Notepad will do).

Thank you for agreeing to participate in this Web Application evaluation. You probably have a good idea of why I asked you here, but let me go over it again briefly. We're asking people to help us evaluate a website to see if the intended feature of refilling reminder button works as intended.

The first thing that I want to make clear right away is that we're testing the web site, not you. You can't do anything wrong here. In fact, this is probably the one place today where you don't have to worry about making mistakes.

If you have any questions as we go along, just ask me. But I want you to try to figure things out as much as you can first. I want to see how you use the application. But if you get stuck, I will assist you.

With your permission, I am going to record what happens on the screen and our conversation. The recording will only be used to help me figure out how to improve the application. It won't be seen by anyone else except the people working on this project. It will also help me because I don't have to take as many notes.

So do I have your permission to record? Thank you.

Note to moderator: When permission is given, start the recording.

- For remote session: Start GoToMeeting screen recording.
- For in-person session: Start Camtasia recording.

Evaluation Proper

- Note to remote moderator: Show the condition 1 / 2 prototype.
- Note to live moderator: Open the condition 1 / 2 prototype.

Note to remote moderator: The instruction below is for the participant to get used to the lag using GoToMeeting.

Note to live moderator: No need for the below.

Before we begin, can I ask you a quick question? How much time do you spend online on a daily basis?

- less than an hour a day
- 1-3 hours a day
- 3-5 hours a day
- 5+ hours a day

Thanks. Now I'm going to give you control of the mouse and keyboard. I'm going to ask that you move your mouse around to get used to the lag but please do not click anything yet.

Start of Scenarios.

I'm going to ask you to perform a specific task.

Note to moderator: Normally a PDF with scenario and task is readied ahead of time for them to read.

Note to moderator: Remember to encourage the participant to try to work through their confusion. If the participant says they are confused for a second time or show frustration, then provide the necessary prompt to the next step they need to complete.

Show the scenario.

Scenario: Create a Medication Refill Reminder

Pretend that you have been recently diagnosed with Type 2 Diabetes. A physician prescribed you Actos 30 mg to help control your blood sugar level. You are using MyMedSchedule.com scheduler to manage your medication.

Task: Create a medication refill reminder for Actos.

Question: What do you think about creating the refill reminder process? How so?

Thank you so much for your participation.

Note to moderator: End the session and stop the recording.

Appendix D: T-Test Output Snapshots

t-Test Results for Time Variables

Student's *t*-Test: Results

The results of an unpaired t-test performed at 15:03 on 8-NOV-2015

t= 2.31
sdev= 70.1
degrees of freedom = 28

The probability of this result, assuming the null hypothesis, is 0.029

Group A: Number of items= 15
28.0 45.0 45.0 61.0 69.0 88.0 99.0 106. 123. 128. 131. 160. 197. 223. 342.

Mean = 123.
95% confidence interval for Mean: 85.93 thru 160.1
Standard Deviation = 82.3
Hi = 342. Low = 28.0
Median = 106.
Average Absolute Deviation from Median = 57.9

Group B: Number of items= 15
20.0 21.0 21.0 21.0 27.0 28.0 38.0 44.0 53.0 54.0 72.0 93.0 102. 167. 199.

Mean = 64.0
95% confidence interval for Mean: 26.93 thru 101.1
Standard Deviation = 55.2
Hi = 199. Low = 20.0
Median = 44.0
Average Absolute Deviation from Median = 37.6

t-Test Results for Error Count

Student's *t*-Test: Results

The results of an unpaired t-test performed at 15:22 on 8-NOV-2015

t= 2.90
sdev= 3.03
degrees of freedom = 28

The probability of this result, assuming the null hypothesis, is 0.007

Group A: Number of items= 15
0.00 0.00 0.00 0.00 1.00 2.00 2.00 2.00 3.00 5.00 7.00 7.00 8.00 10.0 13.0

Mean = 4.00
95% confidence interval for Mean: 2.399 thru 5.601
Standard Deviation = 4.12
Hi = 13.0 Low = 0.00
Median = 2.00
Average Absolute Deviation from Median = 3.20

Group B: Number of items= 15
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00 1.00 2.00 2.00 3.00 3.00

Mean = 0.800
95% confidence interval for Mean: -0.8005 thru 2.401
Standard Deviation = 1.15
Hi = 3.00 Low = 0.00
Median = 0.00
Average Absolute Deviation from Median = 0.800

t-Test Results for Clicks

My Med Schedule: Comparison Study

Student's *t*-Test: Results

The results of an unpaired t-test performed at 15:24 on 8-NOV-2015

t= 2.66
sdev= 2.95
degrees of freedom = 28

The probability of this result, assuming the null hypothesis, is 0.013

Group A: Number of items= 15
2.00 2.00 2.00 2.00 4.00 4.00 4.00 6.00 6.00 6.00 8.00 8.00 9.00 12.0 15.0

Mean = 6.00
95% confidence interval for Mean: 4.440 thru 7.560
Standard Deviation = 3.87
Hi = 15.0 Low = 2.00
Median = 6.00
Average Absolute Deviation from Median = 2.93

Group B: Number of items= 15
2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00 3.00 3.00 4.00 4.00 5.00 5.00 7.00

Mean = 3.13
95% confidence interval for Mean: 1.573 thru 4.694
Standard Deviation = 1.55
Hi = 7.00 Low = 2.00
Median = 2.00
Average Absolute Deviation from Median = 1.13