

SHAUN KELLY | HCI CAPSTONE SPRING 2013

MILESTONE 5 – EVALUATION

Yard Genius Summary

Yard Genius is an application to help homeowners maintain their yards by keeping track of maintenance tasks and suggesting appropriate plants and vegetation based on the user's context.

The main user group for the system will be homeowners with yards and gardens to maintain. Users who will benefit most from the application are those with limited experience and knowledge about yard and garden maintenance.

The three main tasks for the application identified during user research are:

1. Determining suitable plants for a given location
2. Determining which tasks (planting, mowing, fertilizing, pruning, etc.) are required based on what items are planted
3. Design a pleasing layout for plants in the yard

Usability Test Plan

The interactive prototype for the Yard Genius application is functional. The next step is to perform usability testing. This round of usability testing will be formative usability testing. We are testing the application in prototype form to determine what changes need to be made.

Overall objectives for the study

We will gather data about the overall effectiveness of the Yard Genius application. The goals of this study are to:

- To uncover any major usability issues in the prototype.
- To ensure users are able to complete the identified tasks successfully and without errors.
- To determine the user's perception of the quality and ease of use of the application.

Research questions

In addition, in this study will try to answer these questions:

- Are users able to navigate through the application
- Are users able to understand how to operate the search and drag and drop functionality

I collected the following qualitative data:

- The verbal protocol – the running commentary that participants make as they think aloud – will give us indicators about what participants were confused by and why.

I collected the following quantitative data:

- Performance
 - Number of errors in completing each task
 - Task success
- Satisfaction
 - Impression of the usability of the application via the Systems Usability Scale (SUS)

Participants

For moderated usability testing, we recruited four participants (P1: Male, 33; P2: Female, 31; P3: Female, 69; P4: Female, 25). Three are current homeowners. The final participant, though not a homeowner, was responsible for yard maintenance at her home. The participants included all three users interviewed for Milestone 2.

Location and setup

The prototype was loaded for testing in the following location:

superawesomewood.com/ygproto

Participants completed the test in person viewing the prototype on an iPad. Three tests were conducted at the researcher's home. The final test was conducted at the participant's work location.

Participants completed the SUS using Google Forms at the following URL:

<https://docs.google.com/spreadsheets/viewform?formkey=dGJ2WWh5bWthRUxzZ0RHYzN2QlZKQkE6MA>

Methodology

There is a single testing path, so I used a within-subjects design.

The testing session lasted approximately 30-45 minutes, including <5 minutes for test introduction, 20-30 minutes for testing, and 5-10 minutes for debriefing.

The tasks below were chosen to be representative of the types of tasks a user would complete in a given session and to give good coverage to the functionality of the prototype.

Introduction to the session (<5 minutes)

Discuss:

- The purpose of the product

- The purpose of the test; testing the product, not the user
- This is a prototype; not every function works
- Thinking aloud.
- What is your experience using tablet computers? What is your experience using mobile phones or other devices with touch interfaces?

Tasks (25-30 minutes)

Please take a minute to look around the main screen of the application. Tell me what you think the application is for. What tasks do you think you can complete? What are your initial impressions of the usefulness of the application.

Scenario

You used Yard Genius last year to take care of your yard, so you've already added some landscaping and plants. Now it's spring, and it's time to start taking care of your yard.

1. Before you get ready to do your work today, check the weather in Yard Genius.
2. Tell me what tasks you need to complete today and what tasks you need to complete on April 20.
3. Add a new task: Clear branches from yard for the upcoming Sunday.
4. Delete the task on April 13.
5. You've completed all your yard work for today. Mark all of the tasks for today complete.
6. Now that you've completed your to do list for the day, you want to make some changes to your yard. You've decided to add planters to your yard. In Yard Genius, add a planter to the right side of your front walk.
7. You would also like to add some plants. Show me how you would find the plants that require the least amount of care.
8. Add one of the recommended plants to your yard behind your house.
9. You heard about begonias. Determine if they are recommended for your area.
10. Check which items you have on your shopping list.
11. Add a rake to your shopping list.

Post-test debriefing (5-10 minutes)

Discuss:

- What areas did you find particularly challenging or confusing?
- Is there anything you felt like you wish you could do but could not?
- Follow up on any particular problems that came up for the participant.

Survey:

- Ask user to complete the Systems Usability Scale

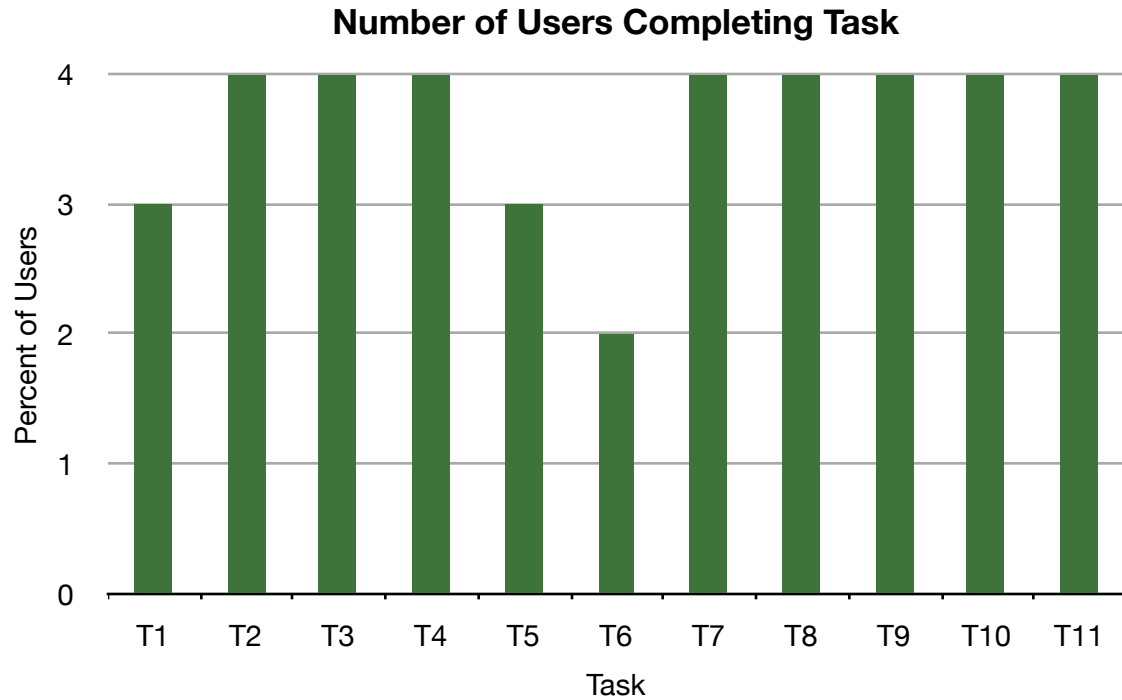
Results

Quantitative Results

Task Success

For each task, participants were scored as either completing or failing to complete the task.

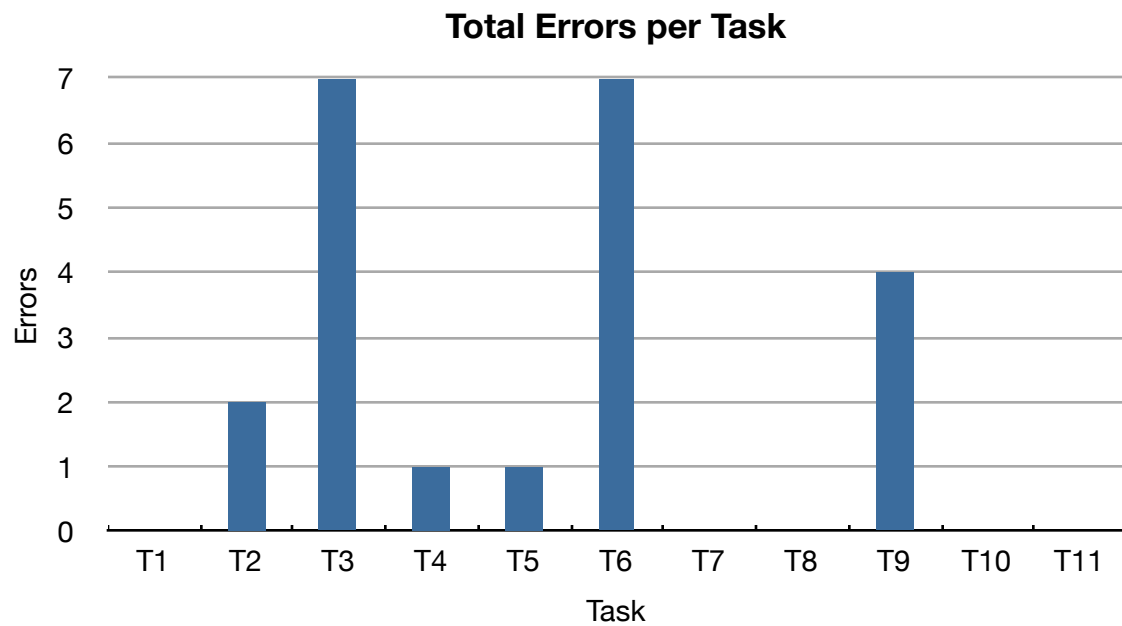
The results for task success are shown below:



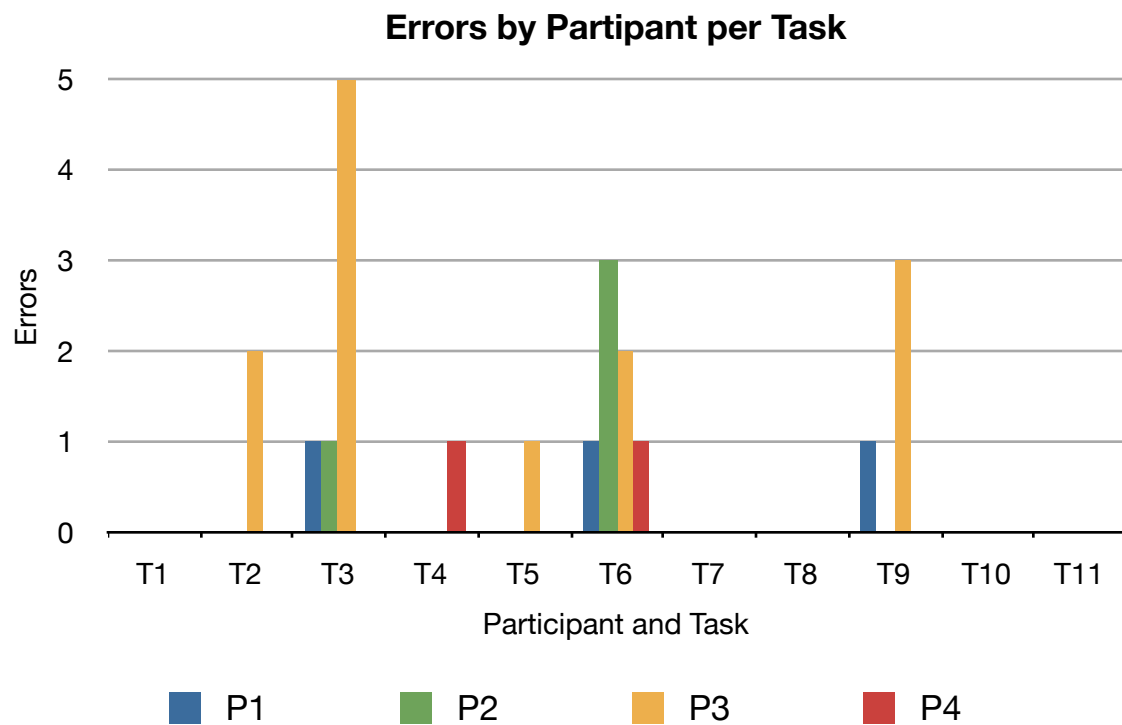
Number of Errors

For each task, participants were scored for the number of errors made when attempting to complete the task.

The total errors by task are shown below:



The number of errors per user per task is shown below:



Systems Usability Scale

Participants were asked to complete the Systems Usability Scale (SUS) after completing the test.

The following table lists the overall SUS score for each participant:

Participant	Score
P1	72.5
P2	87.5
P3	35
P4	82.5

According to research by Sauro and Lewis (2012), when comparing SUS scores between applications, 68 is average and the top 10% of applications score above 80.3.

The average SUS score among the four participants was 69.375 (95% confidence interval :31.6 to 107.15), which is just above average. All but one score was above average and two of the four scores were in the 90th percentile of scores.

It should be noted that the lowest score came from the participant with the least experience with touch interfaces.

Qualitative Results

Major Issues and Themes

The following table describes the major usability issues discovered through testing.

Issue	Severity	Principle	Comment
Task 1			
Home page: Not obvious that you can view weather	3	Visibility	
Task 2			
Calendar Page: Not clear if list on left side is entire list or scrolls to more	3	Visibility, Affordances	Single user issue
Task 3			
Calendar page: Users clicked on date to add new item to date	2	Conceptual model, Affordances	
Calendar obscured by modal dialog for adding new to do item	1	Visibility	
iOS date widget is confusing	4		Single user issue; prototype related
Task 4:			

Issue	Severity	Principle	Comment
Edit and Delete functions are printed light and are far from to do item	3	Visibility, Proximity compatibility	Single user issue
Task 5:			
Calendar page and home page: Users click only on check mark to mark item complete when whole item is touch target	3	Affordances	
Difference between deleting and marking complete not clear	3	Conceptual model	Single user issue
Task 6:			
Landscape page: Users attempted to touch and place before attempting to drag and drop	1	Conceptual model, Affordances	
Not clear of difference between landscape and plants	3	Conceptual model	Single user issue
Task 7:			
Plant page: Not clear if effort means least amount of work	2	Conceptual model	
Not clear if sort by most or least effort	2	Visibility	
Task 9			
Plant page: Not clear what recommended heading means	2	Conceptual model	
Overall:			
Not clear that clear that tasks are inferred by items added	2	Conceptual model	
New feature: Display climate zone on location screen	5		
New feature: Add comparison shopping links to shopping list	5		
New feature: Scan new to do list items for items to add to shopping list	5		

Severity scale:

1. Usability Disaster; must fix
2. Major problem
3. Minor problem
4. Cosmetic problem (more related to prototype itself than system)
5. New feature request

Quotes

Initial Impressions

P4: “Looks like I can make lists, like a to do list or a shopping list. Can probably decide where to put plants in the yard, experiment with their placement.”

Task 1

P1: [On how it detects location] “I assume it updates with GPS as most applications do. It doesn’t show today’s date [w/ the weather], so I don’t know when it was last refreshed.”

P3: [On finding the weather] “Is it even on here? I don’t know how to check the weather...I’m looking for something that says “weather today.”

P4: [On finding the weather] “It had that pinwheel thing that looks like settings. There wasn’t anything else that looked like it might have been weather.”

Task 2

P3: [Beginning the task] “So, I probably want some sort of calendar.”

P1: [On calendar page] “I’m wondering if this is the entire list. Is the list cut off?”

Task 4

P3: “I did it before accidentally, so I knew what to do.”

P4: “I didn’t notice edit and delete because they’re really light.”

Task 5

P2: [On only clicking check mark] “It reminded me of the checkbox in HTML.”

P3: [On clicking Delete to mark complete] “That’s just my guess that’s how you show it complete.”

Task 6

P2: [On why she chose Add Plants to add a planter] “When I think of landscape, I think of big things. It’s not a flower in a pot, it’s the whole bed.”

P1: “Clicking didn’t do it, so I dragged... I assumed after I clicked it, it would highlight so I could see which one I selected and then touch on the screen where I want to put it.”

Task 7

P2: [On the sort order of the filter] “I don’t know if it’s most effort first or least effort first.”

Task 8

P1: “I assumed I could do this the same as edit landscape.”

P3: “I wonder if I can just drag.”

Task 9

P4: [On searching for begonias] “I don’t even know how to spell begonia.”

P1: [On looking for begonias in recommended list] “Don’t know if it means not recommended for area or just not recommended.”

Task 10

P2: [On the sample data in the prototype] “What is ‘pre-emergent’?”

Debriefing

P2: “It wasn’t immediately obvious that you could check weather. It could also tell me if I’m in a particular zone.”

P2: “Did I put all these todos in or did you infer them for me?”

P1: [On additional features he would like to see] “When adding a new 'To do' item, it would be helpful if there is an additional prompt that asks whether or not I also want to add an item to the shopping list to complete the 'To do' task. For instance, the application could pick up on keywords like "plant" or "fertilize" and prompt to add the named plant or fertilizer to the shopping list.”

P4: [On additional features she would like to see] “Find links to anything on your shopping list in case you want to buy online or compare prices”

P3: “I would like it if the program would actually do the job.”

Discussion

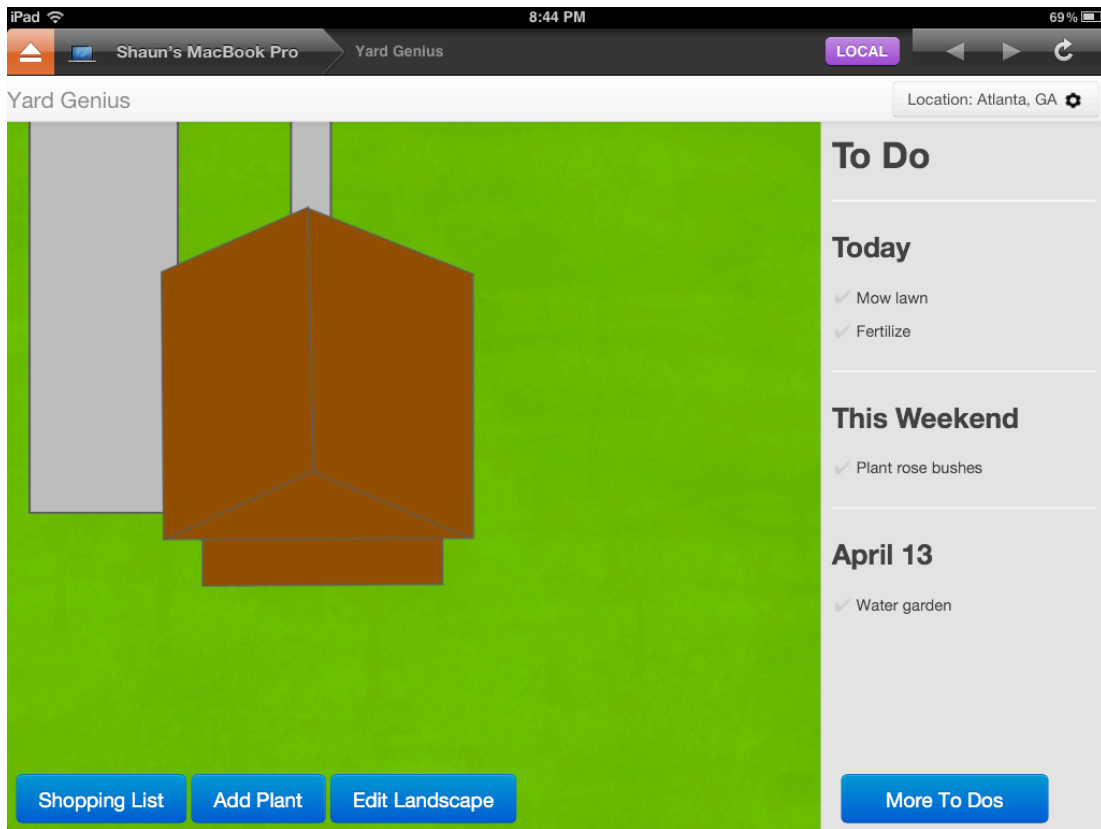
The quantitative data gives a good picture of the overall usability of the system. Users were able to complete most of the tasks in test and were able to complete most of the tasks with few or no errors.

The task-by-task analysis shows that while most of the tasks caused little trouble, a few tasks were more problematic (particularly tasks 1, 3, and 6).

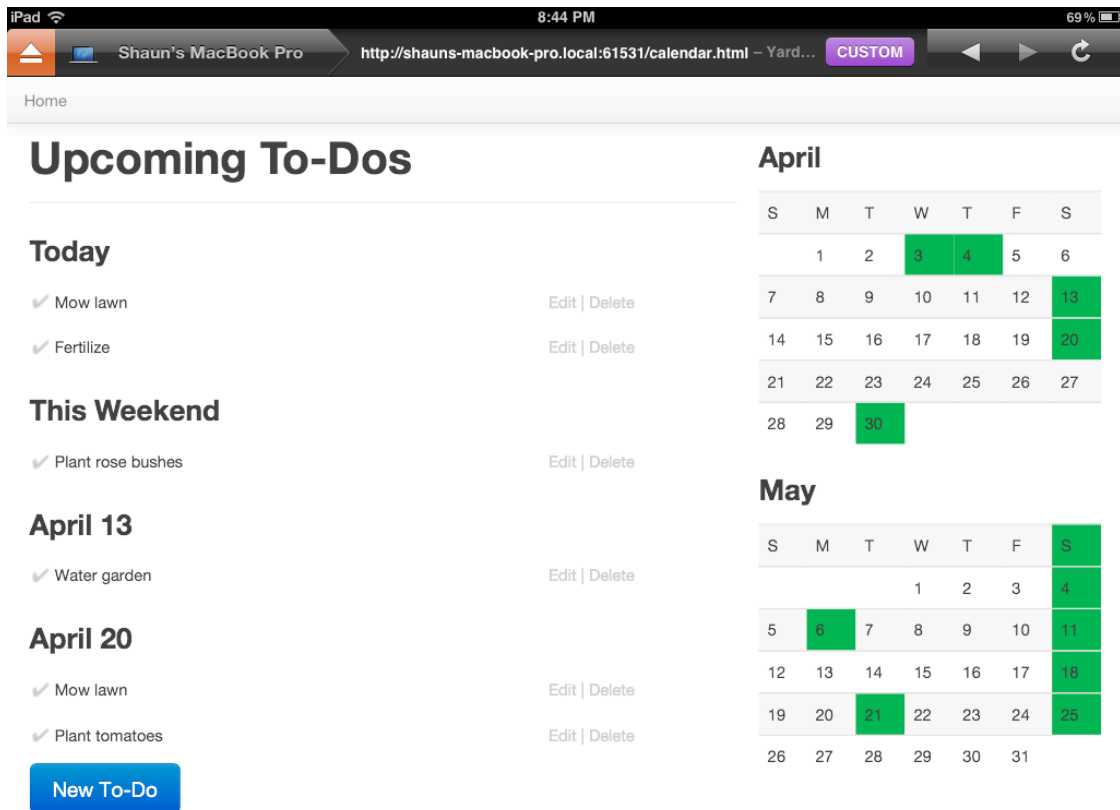
The satisfaction ratings in the SUS scores overall support the picture of an application that is largely easy to use but has several trouble spots.

Again, we note from the quantitative data that one user had more trouble than the others. This user was the one with the least experience using touch interfaces.

By analyzing the qualitative data, we can get a better picture of the specific issues causing the users trouble.

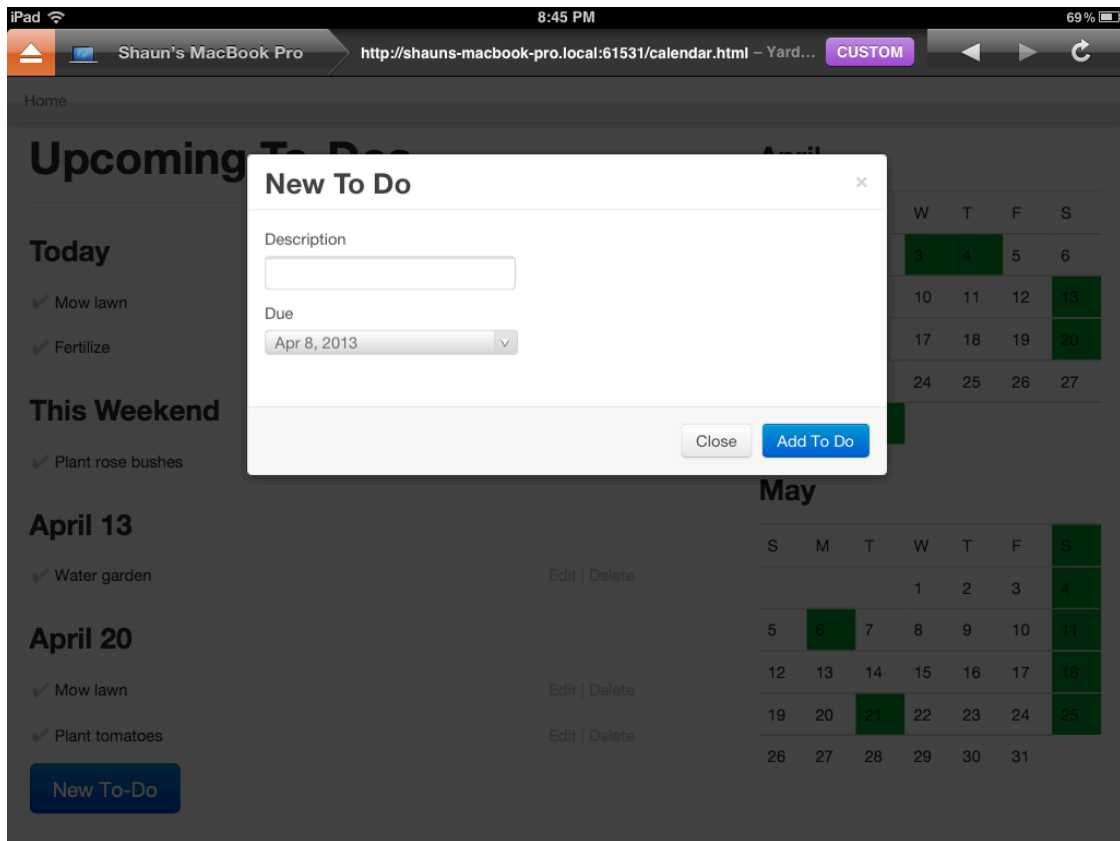


Starting on the home screen (above), we see one of the trouble issues. Users had trouble identifying that the location button in the top right corner offered weather. Had the task not explicitly asked for weather, they most likely never would have looked. This should be updated to provide an indication that the weather function is available.



The More To Dos/Calendar screen (above) had the most issues. First, users immediately tried to interact with the calendar on the right to add new items rather than selecting the New To-Do item button at the bottom of the screen. Interacting with the to do items displayed on the left of the screen also caused problems.

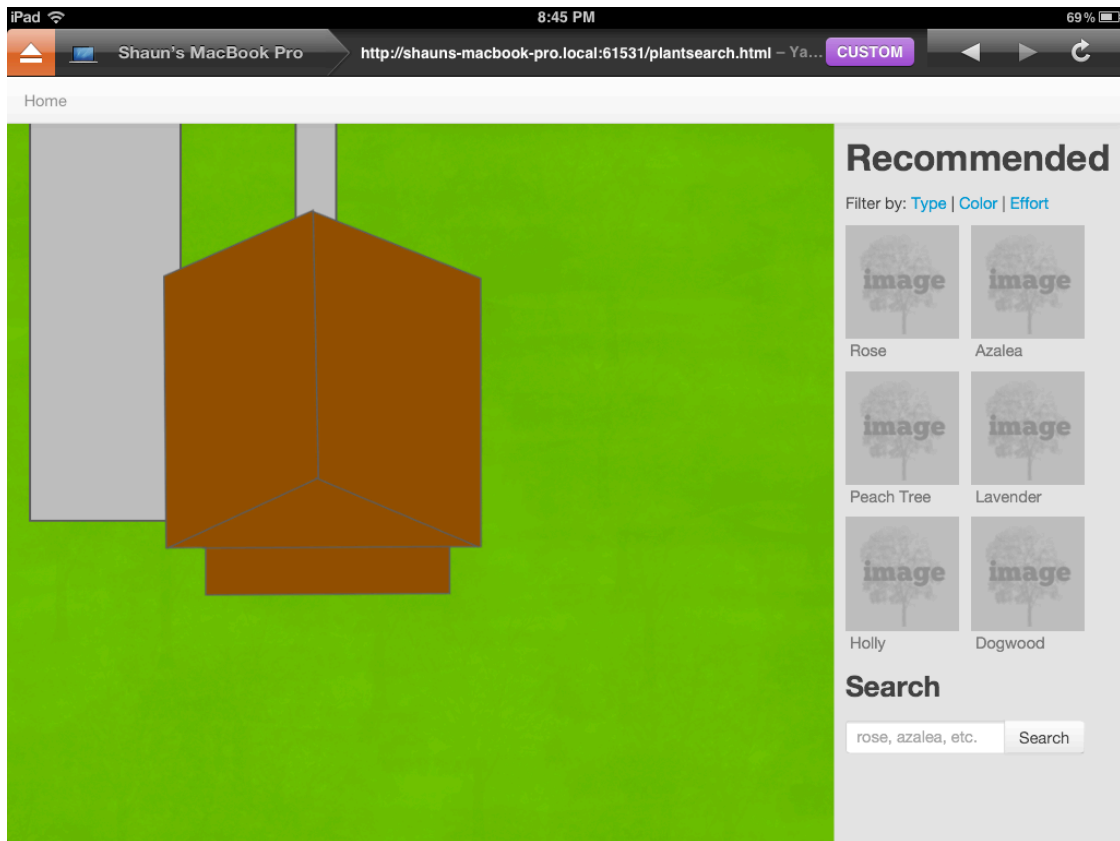
- Users clicked only on the check mark to indicate a task was complete even though the entire line was a touch target. The lines need to provide better affordances.
- The Edit/Delete items were too faint and too far away from the to do items for some users when trying to interact with the to do item.



Additional trouble came when adding a to do item. Users wanted to view the calendars on the previous screen when deciding which date to enter in the Due field. Unfortunately, the modal dialog obscured the calendars (as shown above) and made this very difficult.

I was completely surprised by the issues on these two screens. I had not expected users to want to interact solely with the calendar; however, all four users tested made the same “error.”

While these screens could be tweaked with some simple changes such as moving the modal dialog away from the calendar, I would try to redesign this screen so that the interactions revolve around the calendar instead of the task list on the left.



The final major issue appeared on both on the Add Plant (above) and Edit Landscape screen. These screens were designed to allow users to drag and drop items (plants or landscaping) on to the yard. Drag and drop was important so that users could tap an item in the search results on the right side to view a detail page about the item.

However, all users tested first assumed that they would place items by tapping on the item in the search results on the right then tap in the yard where they wanted to place the item. I would redesign these screens to support this interaction. I would have to add an additional way for users to view plant or landscaping detail pages.

The learnability of this feature was good, though. Users were much better at adding plants using the interaction they learned when attempting to add landscape items.

Users also had a conceptual issue with the application. The major benefit of the application is that it infers which tasks need to be completed by the plants, landscape, location, and weather. However, this was not clear to the users. They did not know whether the tasks in the application had been entered manually or automatically. This issue may not be a problem if users start from scratch with the application, so whether additional changes are necessary requires additional testing.

Reflections

Overall, I found this to be a very effective way to uncover usability issues in the application. While some of the items found were things I had questions about (e.g., the weather issue), others (e.g., the challenges with the calendar page) were things I would

never have found otherwise. This is a good example of the advantage of usability testing over expert review.

Testing a prototype instead of a live application has some challenges. The prototype has to be designed with the needs of the test in mind and the test in turn can only reflect what's available in the prototype. It's also hard to account for all the edge cases and ways a user might attempt to interact with the application, so users might find their way to parts of the prototype that do not work. This is challenging for moderation, though I generally just let the user know that area is not functional and direct them back to functional areas.

One thing I would do differently next time is to measure time on task. I chose not to measure time on task because efficiency is not a main concern for a recreational consumer application and there is some controversy over whether to measure time on task is a valid measure when asking users to think aloud (Rubin and Chisnell, 2008) (Tullis, and Albert, 2008); however, it would provide an additional indication of tasks users found challenging as it is possible to successfully complete a task with no errors even if you spent several minutes searching the screen for an option you just can't find.

Citations

Rubin, Jeffrey and Chisnell, Dana. (2008). *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. Indianapolis, IN: Wiley.

Sauro, Jeff and Lewis James R. (2012) *Quantifying the User Experience: Practical Statistics for User Research*. Burlington, MA: Morgan Kaufman.

Tullis, Tom and Albert, Bill. (2008) *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Burlington, MA: Morgan Kaufman.