Team Optima
AJ Miller
Thomas Curci
Chelky Lin
Shuler Cotton
CPSC 4140
April 24, 2018

**Table of Contents**

## Study Protocol

*Heuristic Evaluations*

Heuristic evaluators used the Heuristic_Eval_Data.xlsx document to fill out their evaluations. These heuristics were directly correlated to Jakob Nielsen's list of 10 heuristics as described in Heuristic_Eval_Data.xlsx. The prototype was also provided. Per the script, the evaluators were asked to test the system before providing feedback in the heuristics sheet. After about 5 minutes, evaluators were asked to begin filling out the sheet. We were in the same room as them the whole time, but were not monitoring their activities. If they had any questions about the interface they would call our name and we would help them. After completing the sheet, we quickly read over what problems they found and asked them for more detail on the problem if it was determined that not enough detail was given. This allowed for a more thorough understanding of the context of their problems so that we could fix it at a later date. After an evaluator was finished, we thanked them and went to the next evaluator.

For our heuristic evaluations, we asked 3 people from our class to evaluate our system. From a chart shown in lecture, we found that 3 heuristic evaluators would find approximately 60% of the problems in our interface. Utilizing more than 3 evaluators would have had diminishing returns on the amount of new problems found. The chart estimates that at 5 evaluators, we only find 75% of the problems in our interface. It was estimated that since we are conducting 3 other types of tests that all together they will catch at least 80% of our problems. The time spent finding 2 more heuristic evaluators and the time spent meeting with them wasn't worth the value that they may have brought to our testing phase.

*Predictive Evaluations: KSLM*

The group decided to conduct predictive evaluations using the Keystroke Level Model (KSLM). KSLM is a way to quantify user actions into time values (in seconds) in order to compare those values to the timed results recorded in usability testing. In order to conduct KSLM for Team Optima's prototype, the team followed each task that was asked of users and recorded the pointer actions, hand movements, button clicks, and keystrokes involved with each of the tasks.

Each of these types of actions has a specific character associated with it:

P - pointing
K - keyboard
H - hand movement
B - button press
D - draw a straight line

After the ideal number of these actions were recorded for each task, the team followed the rules of inserting the “M” operator. The M stands for mentally prepare and is supposed to be added to the actions in the following situations:

1. Insert M operators in front of all K’s that are not part of argument strings.
2. Insert M operators in front of all P’s that select commands.

Then, “M” operators were removed based on the following criteria:

1. Remove M operators that are anticipated by prior operation.
2. Remove M operators if a string of MK’s is a single cognitive unit.
3. Remove M operators with redundant terminators.
4. Remove M operators if the succeeding K is a constant string.

The resulting protocol left the following results for the tasks:

| Task 1 | Task 2 | Task 3 | Task 4 | Task 5 | Task 6 | Task 7 | Task 8 | Task 9 | Task 10 | Task 11 |
|---|---|---|---|---|---|---|---|---|---|---|
| MP | MP | MP | MP | MP | MP | MP | MP | MP | MP | MP |
| B | B | B | B | B | B | B | B | B | B | B |
| MP | MP | MP | MP | | M5K | MP | | MP | M15K | MP |
| B | B | B | B | | MP | B | | B | MP | B |
| MP | MP | MP | K | | B | | | MP | B | MP |
| B | B | B | MP | | | | | B | | B |
| | MP | K | B | | | | | MP | | MP |
| | B | MP | M16K | | | | | B | | B |
| | | B | MP | | | | | | | MP |
| | | MP | B | | | | | | | B |
| | | B | MP | | | | | | | MP |
| | | | B | | | | | | | B |

Finally, each of the operators were summated for each of the tasks and the following time values (seconds) were applied for each of the operators:

P = 1.10
M = 1.35
B = .20
K = .28

For example, the predicted KSLM value for task 1 was 7.95 seconds with the calculation given below:

7.95 = 3*(1.35) + 3*(1.10) + 3*(.20)

This procedure was repeated for each of the tasks and this data was then used for analysis of the task results.

*Think-aloud Testing:*

For the Think-aloud testing, we used our testing script as specified in our Materials Used section. We provided the participants with a computer with the application and the questionnaire open (also specified in the Materials Used section). We then setup an audio recording system to collect the think-aloud feedback from the participants. After an explanation of what the participants will be specifically doing, we had them complete the tasks and answer each questionnaire question following the completion of the task. Throughout the testing, the participants gave audible feedback about their thought mental processes, challenges, and expectations for the system's performance. After completion of the tasks, the audio recording was stopped, the participants completed a demographics section, and then were debriefed.

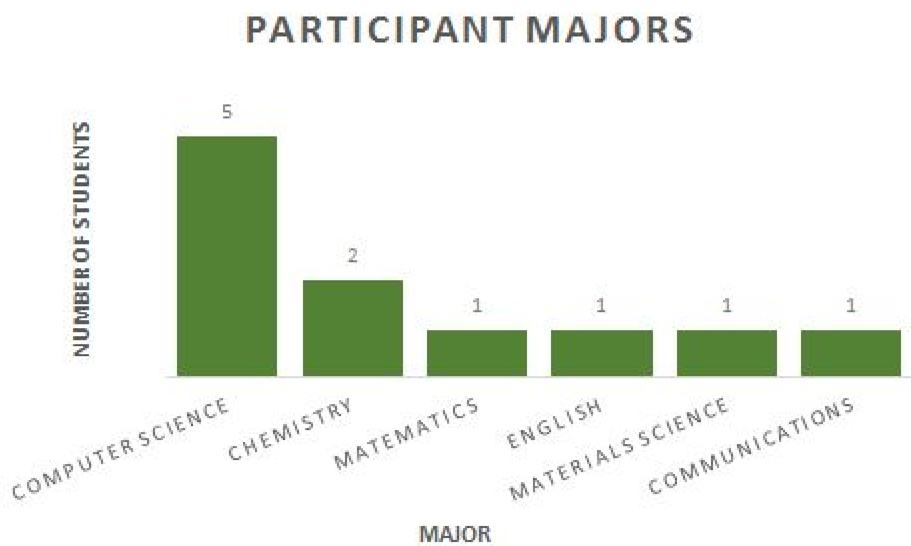
Our five participants were selected from a pool of volunteers who are neither in nor have taken CPSC 4140 and reflected a random sampling as best we could immediately find. The participants were briefed in advance on what they would be completing and how the testing would take place. After accepting that they were willing to participate, we continued with the testing or scheduled a time for them to participate in the testing.

*Usability Testing*

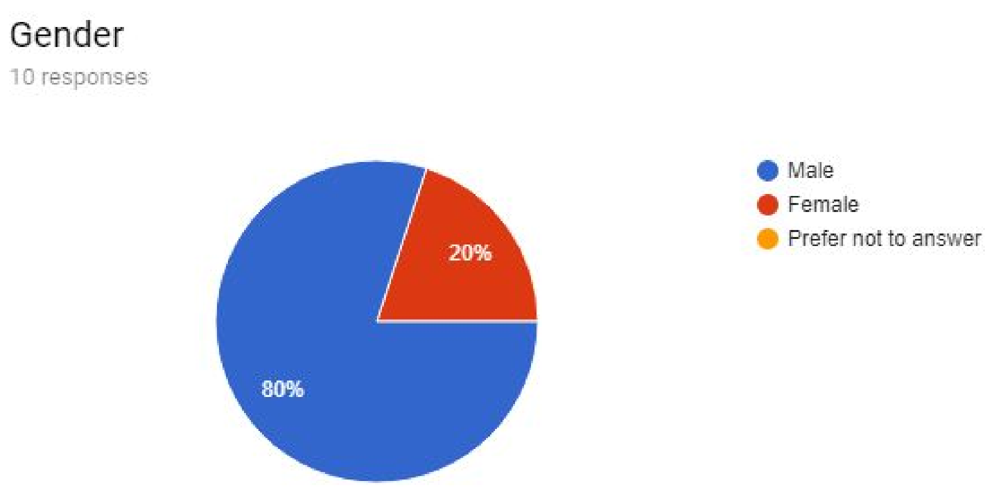
For our usability testing, we ran tests individually and followed our testing script as specified in the Materials Used section. The script introduced who we are, outlined the goal of the study, described what we will do, and asked for permissions. We provided the participants with a computer with the system and questionnaire open. The questionnaire, also specified in the Materials Used section, had tasks and questions about the tasks. The participant was instructed to look at the questionnaire for a task, perform the task on our system, switch back to the questionnaire to answer questions about the task, and repeat for each of the eleven tasks. If the participant took much longer than expected for a task, the participant was allowed to ask for help to complete the task. We recorded the screen as the participants performed the tasks and asked them to fill out a questionnaire during and after the test. The five participants were selected from a pool of volunteers and reflected a random sampling as best we could do.

**Demographics**

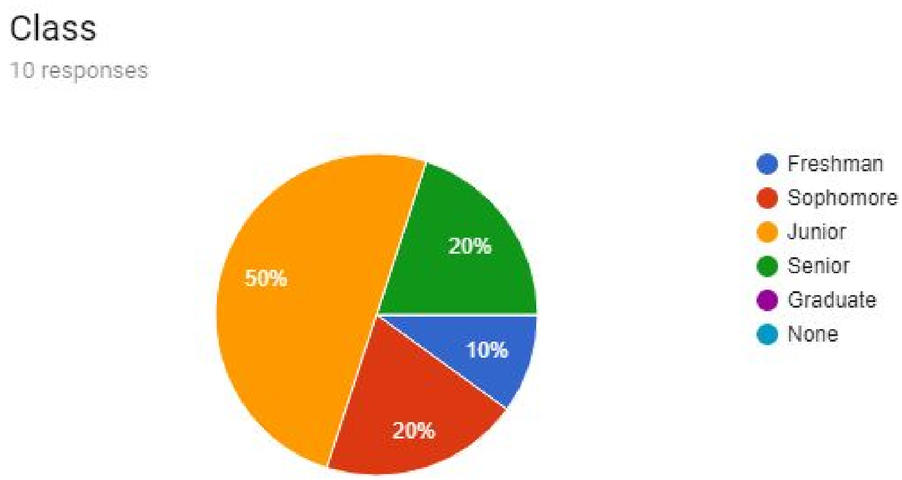
For our demographics we asked our participants 5 questions. The questions were what was their age group, ethnicity, gender, class, and major. For the most part our demographics do not represent our target group which are Clemson University students. We found that 100% of our participants were white and in the age group of 18-24. When doing the usability testing we met with other groups in our class and traded our heuristic evaluations and usability testings. This is why we have an overrepresentation of Computer Science majors in our demographics.




Because Computer Science students are largely represented by males we can also attribute this to our gender distribution which favors males.



Lastly, the distribution of our participants class is highly dependent on who we know. Being seniors, we don't know many freshman or sophomores because we typically don't have classes with them. This is why most of our participants are upperclassmen.




**Description and Justification of Tasks**

Users were asked to complete a series of tasks that allowed the group to test not only the common use cases of the system but also the limit's system and additional features. These tasks are described below, with justification provided for each task.

1. Find the study space that is closest to you.

   *Justification*: This is one of the most common use cases. It was learned in previous milestones that people care a lot about availability and that users will actually go out of their way to go to a space that is generally available. If the user is enabled to find an available space that is closest to them, this is where there study experience begins to be optimized.

2. Find a study space that is silent and has wi-fi.

   *Justification*: These were aspects that were important to users in data collected from previous milestones. "Atmosphere" (50%) and "Noise Level" (75%) were the two most important criteria when respondents were asked, "Which TWO elements of a study/work space are most important to you?" This was combined with users' desire for amenities such as access to the internet.

3. Find a study space that is just for yourself and is quiet within 1 mile of your location.

*Justification*: This task along with task 4 (the next task) was required because it allowed the team to observe how users interact with many different search panes while providing the system with different pieces of search criteria that pertained to their particular needs.

4. Find a study space that is for a group of 3 and is quiet within 16 miles of Anderson County.

   *Justification*: This task is similar to task 3, but with one major difference: to test the function of the “address button” and allow users to search for spaces near a location and not simply based on their current location. Going back to previous milestones, it was important for users to not be limited to their physical location when planning out their work and study spaces.

5. Using the map, select a space nearest the Cooper Library.

   *Justification*: This is a key component to visualizing spaces and selecting a particular space. It is important to get feedback on users completing this task because it has such a prominent display in the UI. If it is unclear how to use or not helpful, then the design needs serious reconsideration.

6. Search for “Watt” in the results search bar.

   *Justification*: Users should be able to swiftly find the search bar and be able to search for a particular location, thus it was important to analyze the user’s ability to do this.

7. Sort a list of study spaces by their rating.

   *Justification*: This enables the team to get feedback on the “Sort By” feature and allows the user to perform the routine scenario which is described in the next task.

8. Select the study space with the greatest rating.

   *Justification*: This is a typical use case in which the user will select the study space that is best for them. “Best” can be described by rating, distance, amenities, and many other factors, but providing some typical ones for them to select from after sorting by was an important task to observe.

9. Rate the selected location.

   *Justification*: Being able to leave feedback on the space and collect data from users that affect the space ratings was a big piece of feedback that was received in the studio sessions. This enables the spaces to be crowdsourced for their volume, comfort, and amenities.

10. Comment on the selected location.

    *Justification*: This is similar to the rating crowdsourcing but allows users to leave feedback and viewing comments is a key component to how users may perceive the space. If users cannot easily leave comments, then this power of the system is rendered useless.

11. Find the street view of the space closest to you.

    *Justification*: This was a boundary scenario in that this might not be used by the user often but is a useful feature of the system and allows the user to visualize the actual space that they are near or are looking for. This will enable them to see the physical locations that they identify on the map.

Overall, these tasks gave the team a firm foundation of user scenarios to test on. Not only were these tasks related to data from earlier milestones, but they also exhausted all of the features of the prototype and made sure to test the usability, functionality, efficiency, and helpfulness of said features. Upon collecting data from users completing the above tasks, the results of their completion were collected and then analyzed for discussion.

## Analysis and Discussion of Results

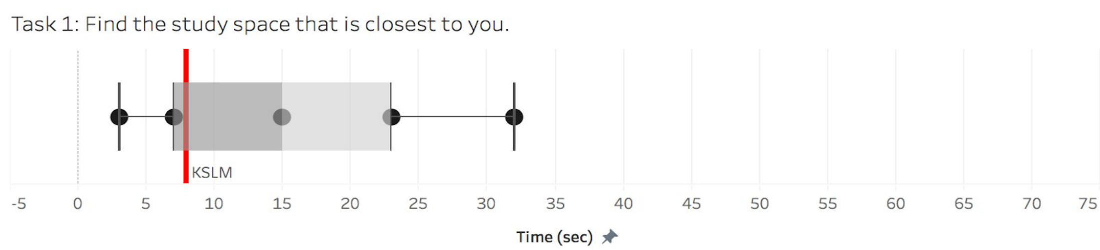
*Predictive Evaluation vs Usability Testing*

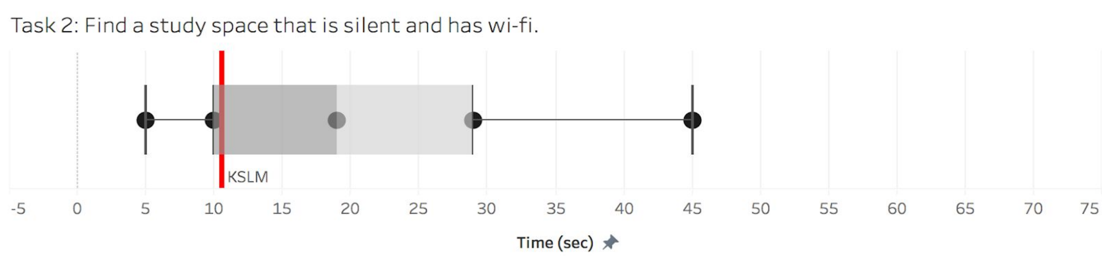
For our predictive model we chose to use Keystroke Level Model (KSLM) to compare the timing of our participants to experimental data collected through usability testing. The below chart describes all of the KSLM values for each of the tasks that users were asked to complete. Each task has a description, a task number (the number in the ordering of tasks that the user completed), the average time, in seconds, taken by users to complete the tasks, the KSLM value for that particular task, as well as the difference between the average actual time and the KSLM value.

| Task Description | Task | Time (sec) | KSLM | +/- |
|---|---|---|---|---|
| Using the map, select a space nearest the Cooper Library. | 5 | 47.20 | 2.65 | 44.55 |
| Find a study space that is for a group of 3 and is quiet within 16 miles of Anderson County. | 4 | 41.40 | 19.36 | 22.04 |
| Search for "Watt" in the results search bar. | 6 | 27.40 | 8.05 | 19.35 |
| Sort a list of study spaces by their rating. | 7 | 22.00 | 5.30 | 16.70 |
| Find a study space that is silent and has wi-fi. | 2 | 21.60 | 10.60 | 11.00 |
| Find a study space that is just for yourself and is quiet within 1 mile of your location. | 3 | 24.20 | 13.53 | 10.67 |
| Find the study space that is closest to you. | 1 | 16.00 | 7.95 | 8.05 |
| Select the study space with the greatest rating. | 8 | 9.25 | 2.65 | 6.60 |
| Comment on the selected location. | 10 | 14.00 | 10.85 | 3.15 |
| Find the street view of the space closest to you. (Find Edwards) | 11 | 10.50 | 13.25 | -2.75 |
| Rate the selected location. | 9 | 6.67 | 10.60 | -3.93 |

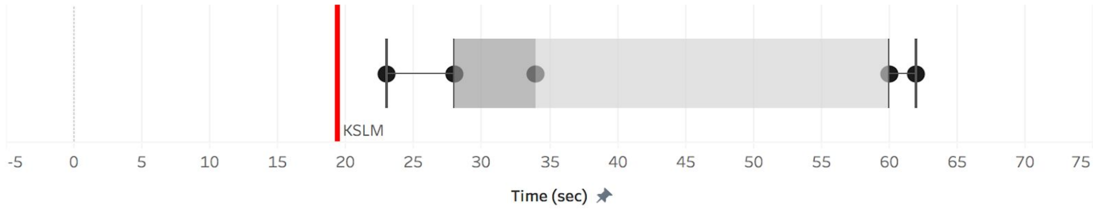
There seems to be a trend that earlier tasks took longer than later tasks even if we thought of them to be of the same difficulty. The 5 largest KSLM differentials (meaning the slowest tasks) were averaged at task number 3.8. The 5 lowest KSLM differentials (meaning the fastest tasks) were averaged at task number 8. This shows a clear bias in terms of the earlier tasks being slower because the user was still learning how the system generally worked. After the user gained experience in the 1st half of tasks, they were able to complete the 2nd half of tasks much quicker. If we were to conduct these evaluations over again we would like to add a few tasks at the beginning that navigate the user through the system without collecting data.

In the next figure, each of the tasks' distribution of times are depicted in the gray box plots while also highlighting the KSLM value with a red line for each plot.

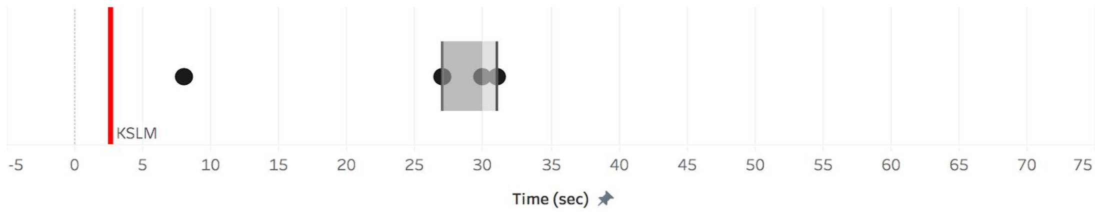


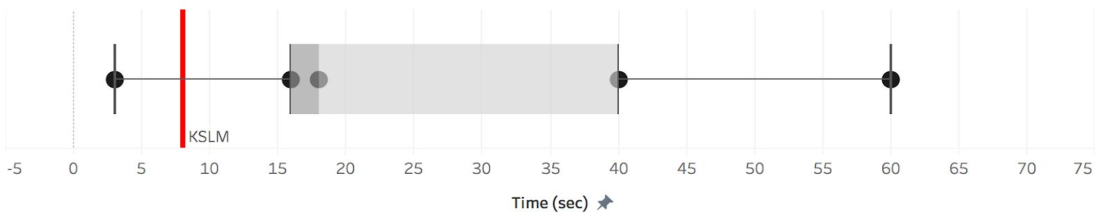

Task 4: Find a study space that is for a group of 3 and is quiet within 16 miles of Anderson County.



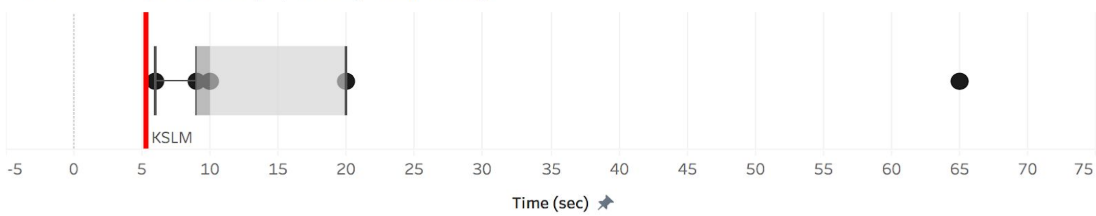

Task 5: Using the map, select a space nearest the Cooper Library.



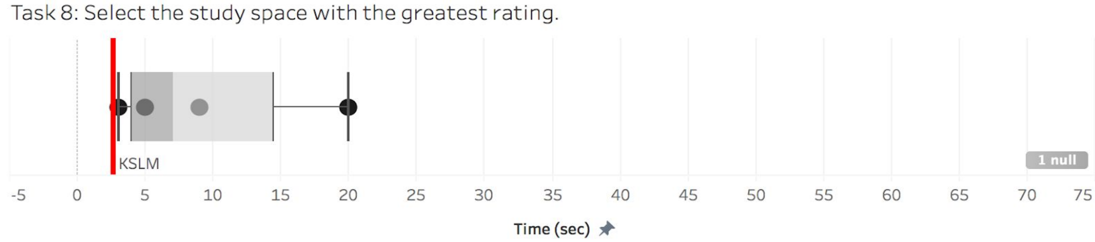

Task 6: Search for "Watt" in the results search bar.



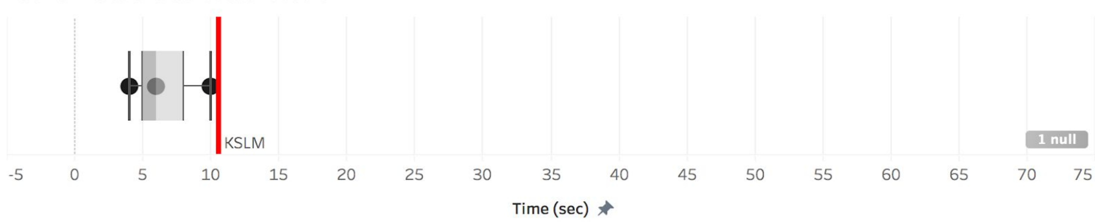

Task 7: Sort a list of study spaces by their rating.



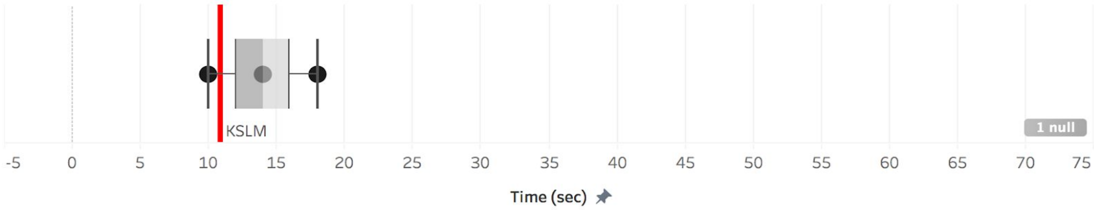

Task 8: Select the study space with the greatest rating.



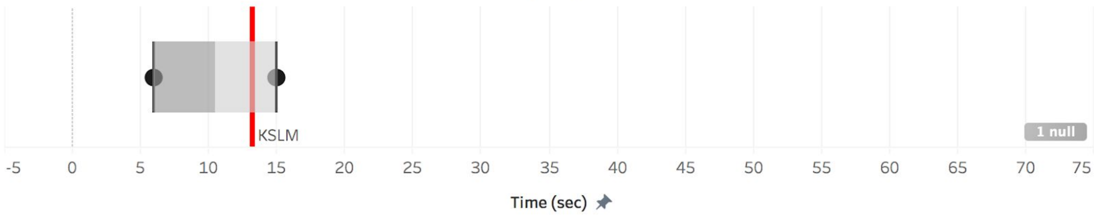

Task 9: Rate the selected location.



Task 10: Comment on the selected location.



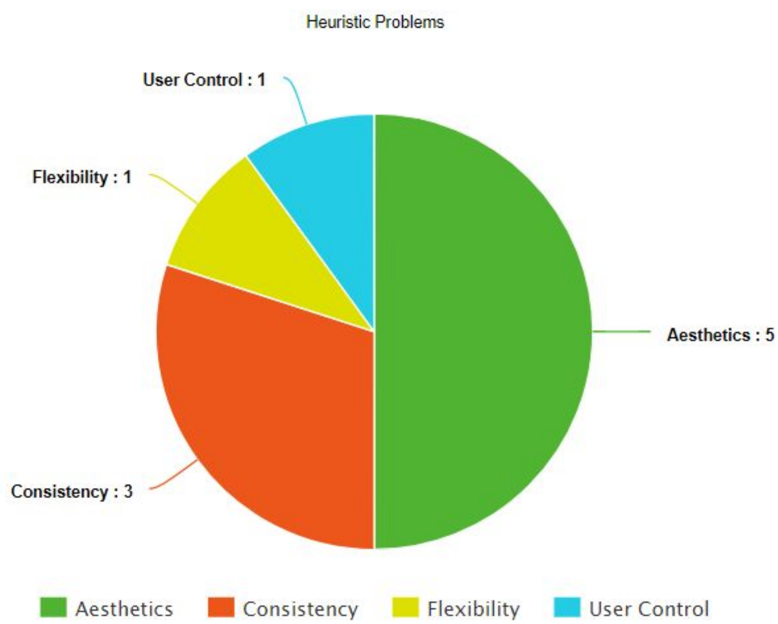

Task 11: Find the street view of the space closest to you. (Find Edwards)




This figure highlights an explanation for task 5 being the greatest difference between actual and predictive times: the median time for the task was not that different from many of the other tasks but the KSLM value is extremely low and there was an outlier that significantly increased the average. Additionally, there were some participants that went off task using the map. There were also some students who were not familiar with the campus layout so they took a large amount of time to complete the task.Many of the KSLM predictions fell on the lower range of the actual times recorded. There are two potential explanations the group has found: inaccuracies of the KSLM model and usability flaws in the interface. These flaws are explored later in the results of the think-aloud evaluations and heuristic evaluations.

*Heuristic Evaluation*

Our 3 heuristic evaluators found a total of 10 problems with our website. Of these 10, there were a couple of problems that were discussed across multiple heuristics. When we discussed the problems mentioned as a team we found that a few of them were negligible because the user didn't have to use the feature to use the application to its fullest potential. As an example, some of the evaluators said that we should remove street view because it's not necessary.  When we added Google Map API, however, the API added this feature in there for "free" and removing it would not benefit any of our users.




On the other hand, there were some problems that we agreed with and would like to fix. A few of the problems that our evaluators discussed were seen in our usability testing as well. They found that the "X" button to minimize the information on a recently clicked building was not consistent with the action we typically expect an "X" button to do. This was shown to be the case in the usability testing because when participants were trying to get back to the preferences screen they didn't think to click that "X" button at first glance, they only thought to do it after process of elimination, which took much longer. That button should be made more clear as to what it does.

The other common problem that our evaluators found was that when you are browsing the pictures for a building, the horizontal scroll bar is not intuitive to use. Instead, they suggested we use a button to switch between the pictures. Our team agreed that using an arrow system to go back and forth between the pictures would both look better aesthetically and feel better functionally.

The mean priority across all of the problems our heuristic evaluators found was 3.8 on a scale of 4.0 where 4.0 is the least priority. For our first round of testing I think we created a pretty good interface for our users. Most of the problems found were trivial, the others were good catches and would be fixed in the next update of our website.

*Think-Aloud Testing*

The Think-aloud testing gave us a substantial amount of feedback involving general bugs, confusing design choices, and features users would like have to that the prototype did not include. Overall the suggestions were very helpful in guiding us to see what the most crucial functions of the system are and what users with a brand new experience with the prototype are thinking and seeing when they try to use the system. The feedback and observations we received from them are discussed below.

Multiple participants were confused initially because the system did not indicate that it was starting with their current location and they often believed they needed to complete some steps to initialize the system. There were also two instances of confusion involving finding study spaces on the map when they weren't actually displayed in view. We could resolve these two issues by relaying to the users that when the system starts it marks your current location when the scope zooms on it. We could also have the system automatically zoom to include the most nearby study spaces, so they can immediately see roughly how close/far they are and where they may begin their search visually.

Another minor design flaw that caused confusion was the Find My Location button not being as prominent as it should be since it is the primary submit button within the system, but is a grayed out color while the my location is bright orange. The find my space button will become bright when hovered over, but that indication comes after the user has already found it. We should consider some changes to the left side pane to consolidate the options including making all of the options encapsulated into drop down menus that can be skipped if the user doesn't care and can be unselected if they want it to go back to blank, which was a common comment. We should display the find my space button with greater visual prominence to indicate the user should select it to narrow the scope of their search.

It was also requested that we include a reset button next to the find my space button, so that users can easily clear their entire selection and have the scope narrowing reverted back to no selections to see all of their nearby options. It also wasn't immediately clear to all users that they needed to press the find my space button following choosing options rather than the system just automatically updating when the selections in the left pane are made. This is an interesting idea that would need more discussion put into the design decision, but the system could potentially not have a find my space button at all if it automatically loaded once the selections each are made. This may become confusing and cause the system speed issues if it gets hung making a new calculation for every selection, but is definitely a consideration to be made.

We also found that users tended to struggle with entering a custom location because the menu didn't make it fully clear that they could enter anything they wanted and not an entirely correct address such as just the phrase "Anderson County." After a new address is entered, the system should also change the view to that location and zoom out per our prior observation. Another challenge with this part of the system was that users struggled with typos and with ambiguous entries such as "Cooper Library" because there are more than just the one on Clemson's campus that they were looking for. It would be helpful to change the name of the feature to custom location and if users enter data into the field that we call on Google Maps API to return suggestions for what they are looking for that they are able to select and include distance from the current location to help with their choice.

A minor comment we received from a user was that it was not clear when a rating was submitted that it was actually accepted because the submit button just changes the ratings made and turns it into the average ratings. It would be more clear if there were an indication of success upon submitting a rating that we would add in another iteration of the prototype. It was additionally not clear in the list of locations which ones all met the criteria that was included on the side pane such as noise level and amenities (include the amenity icons when you select them on the left). We were also asked to add some sort of popup on the map to show which location is which when hovered over or some sort of more subtle indication of which locations are which on the map and that the pins are study spaces.

Other bugs and small issues we found included being able to place street view outside of the map view, searching for a specific location in the list causes the other options to disappear and not return upon clearing the search, there not being a clear option on the search, locations that have been sorted out by not meeting selection criteria still show up on the map, selecting a location in the menu does not indicate it on the map, right clicking for street view does not work, sometimes the current location cannot be found by the browser on refresh, when the current location is working the left pane gets stuck on the selection part and inhibits the use of the system, there is no way to cancel searching for current location, the ratings on the bottom pane are not clear they are ratings, don't use alerts instead of a notification in the module for issues/errors, fix the favicon, and make all of the panes appear more consistent.

*Questionnaire Results*

From the questionnaire results, a large majority of the participants found all the tasks very easy and did not have problems completing the tasks. A few of the outliers can be attributed to the participant creating an error state that was unrecoverable or difficult to recover from, short of refreshing the page. When considering the participants' responses, we noticed that their responses were mostly consistent with how quickly they performed the task. This shows that the system was mostly intuitive to use when trying to perform the tasks we specified. The participants' responses to "The system was easy to use" was nearly equally split with one more vote for "Most Important" compared to "Neutral" and "Least Important." The responses to "The system was visually appealing" was mostly "Least Important" with one more vote than "Neutral" and only one vote for "Most Important." The responses to "This system worked quickly" was mostly "Neutral" with one more vote than "Most Important" and no votes for "Least important." Overall, from these ranking questions, the general consensus is that ease of use is moderately important, the speed of the system is not unimportant, and the appearance of the system is least important. However, this is misleading because while the participants may not actively care for the aesthetics of the system, general knowledge of interface design tells us that users do care about aesthetics and it should not be ignored. We also asked participants “How often would you use an application like this to find study spaces?” and half of the participants responded “Sometimes,” four of the ten responded “Rarely,” and only one participant responded with “Often.” There were no votes for “Never.” This shows that while the system may be used rarely, it is still useful. However, this also shows that the system does not yet have features that pull the participants to say that it is so useful, they will use it often. We do have to keep in mind that this question is subjective and depends on the each participant’s concept of each answer. We read the responses to the open-ended feedback questions and discussed them in earlier sections. We also discuss responses to participants’ suggestions in the upcoming Future Work section.

**Future Work**

If we had another semester to complete a second cycle of User-Centered Design, there are several areas we would do differently or improve. First, we would begin by completing another Milestone 1, particularly to improve user requirements analysis and gather better and more relevant info about user needs. With this information, we can redesign and improve the system accordingly. We would choose a design framework to assist with layout, animations, and fallback support rather than writing almost everything from scratch. We would also consider developing design guidelines or adapting an existing set of guidelines. Another improvement is to rewrite the code to allow for functional and stateless programming. This would allow for a central managed state with few or no side effects. We would also improve upon Milestone 4. We would consider the Hawthorne effect and allow participants to test independently in a room without an observer. In some of the usability tests, we also found that participants could reach error states that they could not recover from which could affect some of our results. We would also consider adding dummy tasks that would introduce the participant to the system and not count the data gathered for those tasks. There are also a great deal of other improvements. The system would benefit from showing capacity and floors with room layouts. Some usability testers mentioned that there could be more sort options such as alphabetical, noise level, and availability. In some usability tests, we noticed that the participants could benefit from a method to reset inputs quickly so we would add a way to reset all selected criteria and clear all entered inputs. Observing the participants also showed that they expected typed input to be automatically accepted rather than having to hit the enter key to submit the input. To address this and improve user flow, we would filter the results as the user provides inputs and selects criteria. In considering performance, we would give the user a delay before filtering to reduce load on the back end and prevent undesired changes. Related to this, the map could pan to the input address and zoom in or out according to the nearest results. To address user control, we would give them the option to lock onto the GPS location instead. Also related to the map, the system should hide elements that are not related to the results but are included in the Google Maps API such as restaurants or shopping centers. Another participant mentioned that legibility was an issue in some areas of the system. We would improve legibility by adjusting colors, font size, emphasis, and grouping of elements. We would also improve terminology which The same participant pointed out design inconsistencies, particularly in the button stylings between rounded, square, and colored with drop shadow stylings. In one of the studios, one suggestion we received was the ability to see the ratings as a function over time so we would display a graph that shows how the various ratings fluctuate over the course of a day, a week, or even a year. Other improvements include error handling. Rather than using browser alerts, the system would use custom modals to avoid interrupting the user flow. We would also consider adding a splash screen with tutorials to introduce users to the system and provide usage examples. Lastly in this list, we would improve the professional aesthetic by finalizing the system name, choosing a logo, and adding a favicon to the site. These items are what we would change and improve in a second cycle.

### Materials Used

Heuristic_Script.pdf
Questionnaire.pdf
Thinkaloud_Data.pdf
Thinkaloud_Script.pdf
Usability_Script.pdf
Heuristic_Eval_Data.xlsx
Usability_Testing_Data.xlsx