

12 HOURS/WEEK

NILA BANERJEE

PROJECT 6

USER-CENTERED RESEARCH & EVALUATION

APRIL 2017

WHAT IS THIS?

Spring of my junior year, I signed up to take a class called 'User-Centered Research & Evaluation,' mainly because it was required for my major in Human-Computer Interaction, but also because I was curious and excited to learn more about user research. I'd learned about a few techniques at a high level in my previous classes, but this course promised to "educate students from all backgrounds about the human-centered research process and evaluation techniques" in "a learn-by-doing environment." Essentially, I'd signed up to spend 12 (usually more) hours of each week finding out how to best ~~stalk~~ research and understand people and try to convert their needs and desires into ~~an~~ human-focused solutions.

The following blog/memoir/report is divided into three parts and covers my time in this course:

Part I: "The Big One" An 8-week long research project conducted by my team looking into Pittsburgh's housing crisis.

Part II: "The Explorations" Three short projects using three different evaluation methods – Think Aloud, A/B Testing, & Heuristic Evaluation.

Part III: "UCRE & ME" What I got out of this course (besides a resume boost).

PART I: THE BIG ONE

Applying user research methods, team ICAN investigated the Housing Choice Voucher Program, made to combat the affordable housing crisis in Pittsburgh, and developed insightful findings that informed preliminary solutions to alleviate related issues.

Team: Imogen Todd, César Neri, Anqi Wan, Nila Banerjee

Timeframe: 8 Weeks

Deliverables: Oral Presentation & Formal Report

On the first day of class, we found out that not only would we be working on a half semester-long project, but that it would begin immediately, before we received any significant background in research. I was placed in a group with 3 amazing people and began a journey that culminated in a 10-minute presentation and a 63-page process document. Our work was also presented to the mayor of Pittsburgh, as well as community members.



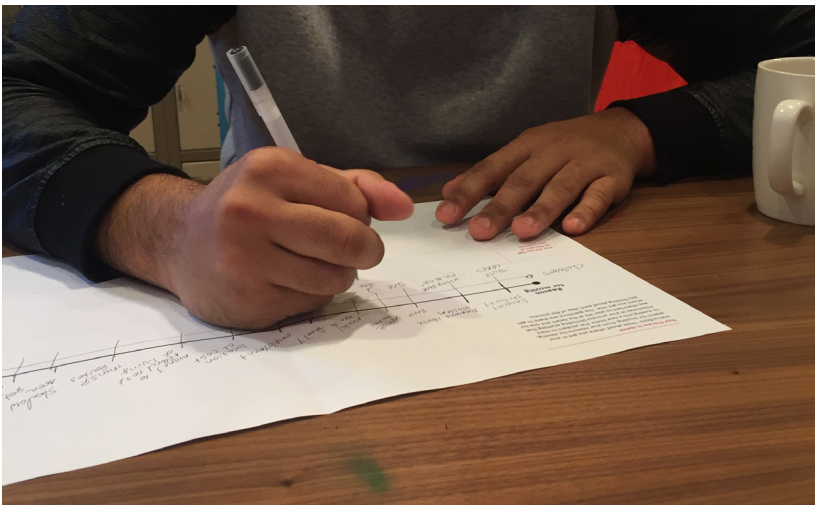
Research

Team ICAN's mission was to *facilitate the transition into affordable housing for low-income residents*.

We specifically chose to focus on the Section 8 program as it is the nation's largest rental assistance program and one of the most popular avenues sought out by those in need of housing. We began by conducting domain research and interviewing a motley of stakeholders in this issue: Section 8 applicants, landlords, social workers, and government officials. Conducting these contextual inquiries helped us better understand the perspectives of these diverse groups, while experiencing their environment gave us a deeper level of empathy for them. I conducted multiple interviews, diagrammed data, and helped brainstorm and format insights.

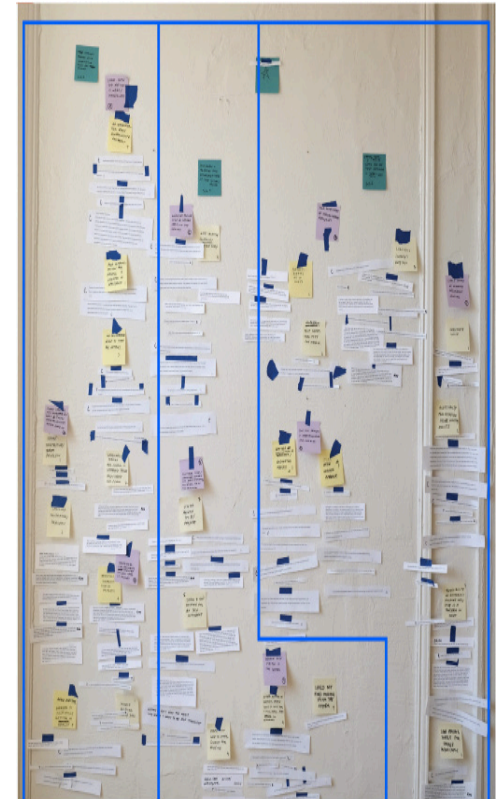
From our research, we identified 3 key findings:

1. Individuals & agencies hold *discordant views* of the system and the process.
2. The *channels* through which information flows *are unreliable*.
3. The *supply* of housing, information, and resources often *does not meet the existing demand* in Pittsburgh.



An interviewee filling out a timeline of his housing application

Our affinity diagram broken into the key findings



Furthermore, we conducted a survey on Pittsburgh residents and found that *50% of respondents were not aware of the local housing crisis* and most of them were prevented from donating or volunteering to charity-minded organizations due to financial or time commitment reasons. These results demonstrate a *need for increased contact between the organizations supporting low-income residents to the larger Pittsburgh community* to foster greater engagement. I was in charge of creating and disseminating the survey and coding the data from it. From this research, we created several models and diagrams to better understand the various relationships, factors, and groups at play in the housing crisis. These are highlighted on the next page.

Methods & Models

We used a number of methods & models over the course of the project. I describe some of them here.

Stakeholder Map: diagram the influences surrounding an issue/demographic (placed at the center) with more important factors closer to the middle. We modeled the stakeholders related to residents in need of affordable housing, which include employers, family, local businesses, and non-profits.

Flow Model: demonstrates the relationships of an individual, highlighting bridges and breakdowns. These were drawn for each interviewee and helped us analyze relationships and discover patterns of breakdowns, including a lack of strong familial ties in most interviewees.

Persona: a depiction of an individual who represents a key demographic, illustrating their goals, frustrations, motivations, personality, and more. Based on our interviews, our personas modeled a male in need of affordable housing and a female case worker.

Contextual Inquiry: conducting an interview and observing someone in context which we are designing for to develop a better understanding of who the user is and where/how they work. We conducted CIs at non-profits, government offices, and apartment buildings when talking to our interviewees.

Cultural Map: a diagram of the groups and communities with which an individual identifies, including family, friends, workplace, location, and more. We notice that the non-profits tended to fill a space usually occupied by family and friends in our Section 8 applicants.

Storyboard: illustrates a scenario that demonstrates the application of a solution using a user's 'see, think, do' perspective. We created storyboards that showed how a subsidized housing applicant's time could be made more productive and encourage self-sufficiency.

Affinity Diagram: a way of breaking up interview data into different groups and subgroups to detect families of ideas. We spent about 3 hours creating an affinity diagram to synthesize our research and discover our key findings.

Sequence Model: a timeline depicting relevant events in an individual's life. These helped us understand how people came to be in need of affordable housing and what they did while waiting for their application to be processed.

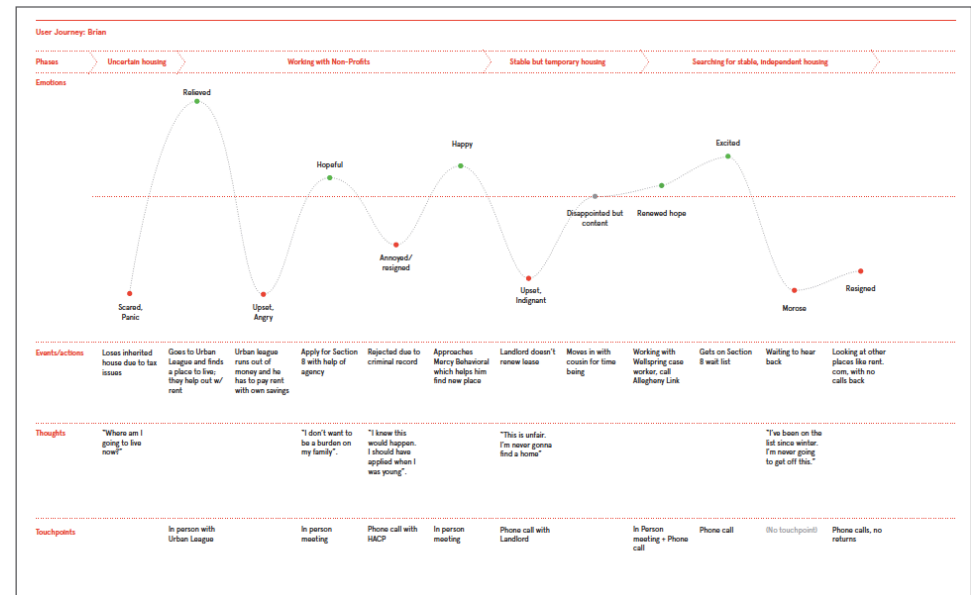
These tools helped our team dig deeper into our research and identify patterns, analyze statements, and discover opportunities for improvement. I participated in the development and execution of all of these models, drawing multiple diagrams and working with my team to produce and refine findings.

Findings

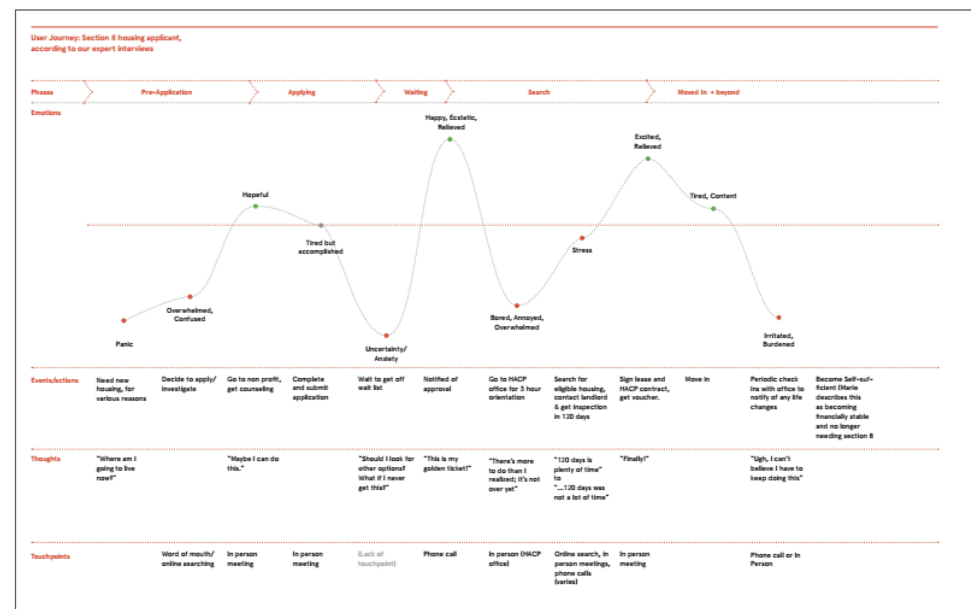
After a lot of drawing, brainstorming, and analyzing, we were able to accurately grasp the difficulties people faced in the Section 8 process. Integrating what we learned from the diagrams drove us to not only empathize with low-income residents, but also uncover opportunities for improvement. In particular, we used the journey maps to identify the following major problem areas:

1. People *lack information* during the “waiting” phase.
2. There is *no centralized list* of all affordable housing resources.

Our final solutions focused on uplifting people by making their lives more productive. Even if they are unable to obtain housing for some time, low-income residents should be connected to programs that will allow them to improve their job prospects, stabilize their mental health, and make plans for the future.



A typical experience in the current system



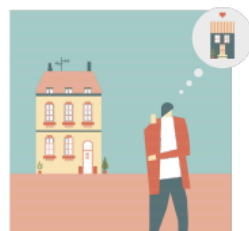
The experience we want to create

Future Directions

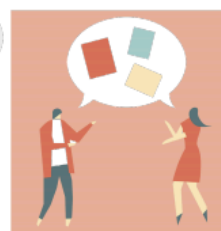
If we had more time to work on this project, we would start by fleshing out our solutions in greater detail and getting feedback on them from the community. By taking a social design approach, we would be able to gain the perspective of stakeholders on the solutions and redefine them to better suit the needs of the individuals affected. Moreover, we would have gotten in touch with other community members to better assess the numerous factors that influence the Section 8 housing program. Finally, we would have looked into other cities who have faced similar issues and used their solutions to inform our proposals.

As our research currently stands, the greatest priority for the city seems to be the lack of housing and employment options for low-income residents. Working to ameliorate both of these would help Pittsburgh ensure that its residents are not forced onto the streets.

How might we modify the Section 8 process to promote or encourage a higher level of self-sufficiency?



Teddy has aged out of the foster system. He needs to find a home and doesn't want to rely exclusively on the government.



He applies for the self-sufficiency program, and his case worker connects him to various resources.



Six months later, Teddy has a stable job and is able to pay 50% of the rent in his new apartment.



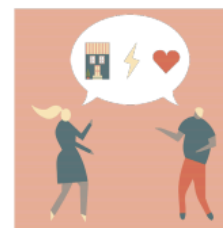
Teddy graduates from the program with help of his case worker.

Solution 1: fostering independence among applicants

How might the HACP make time spent on the waitlist more purposeful and less aimless for younger applicants?



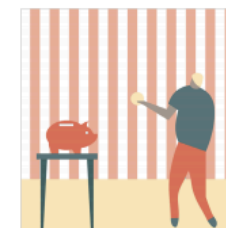
Sean has applied for the Section 8 Voucher Program and is now on the waitlist. He's looking to move out of his cousin's apartment.



Sean talks with his caseworker to formulate long-term goals.



His caseworker gives him updates about his housing application as well as resources to realize his goals.



A year later, Sean is still on the waitlist, but is now saving up money to go to trade school and become an electrician.

Solution 2: ensuring productivity among applicants

PART II:

THE EXPLORATIONS

Over the second half of the semester, we explored a number of different usability methods and conduct shorter individual projects to apply these methods. Here, I explore three of my projects in detail.

Think Alouds

Team: Nila Banerjee

Timeframe: 1 Week

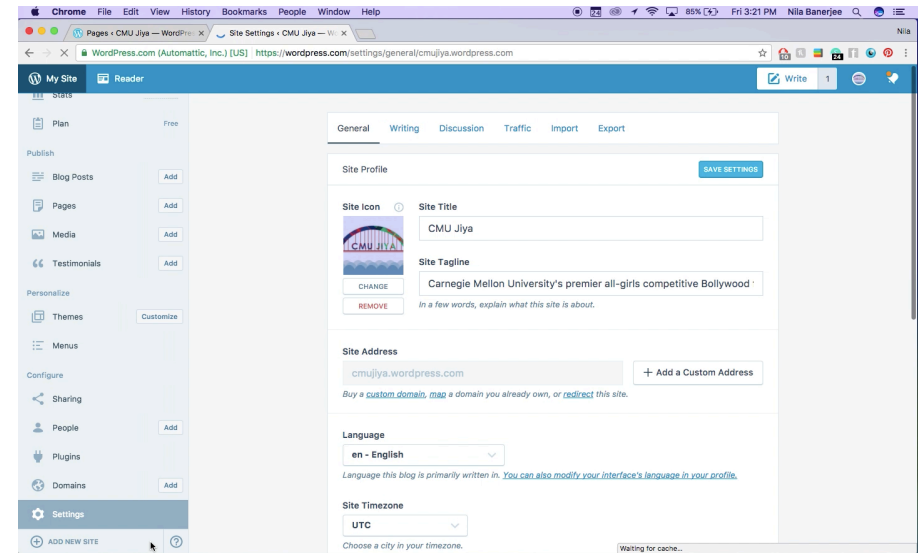
Deliverables: Formal Report with Findings

In one of my favorite projects, I conducted think alouds on WordPress, a popular service for website hosting and management. Think alouds are performed by presenting a participant with a product or service and asking them to complete a set of pre-defined tasks, while they talk through what they are doing. This is an especially powerful method of user testing, as it's the closest we can get to actually *being in the user's brain*. As users work through a task and narrate their journey, the researcher gains a better *understanding of what the users are actually thinking and feeling*. Think alouds are also quite successful as research shows that a large majority (around 80%) of usability problems can be identified with less than 10 tests. They can also identify problems that may have been hidden due to expert blindness, or the knowledge that comes with familiarity. Finally, think alouds are usually relatively *fast and low-budget*, making them a flexible option to test many products and services.

To perform a think aloud, first, gather your participants and brief them on what they'll need to do. Explain and demonstrate what a think aloud is before asking them to do one. The most important thing to keep reiterating is that the test is of the product, not the person. Users may begin to feel frustrated or confused; remind them that it's most likely due to the product's shortcomings and ask them to point out where, why, and how they had a negative experience. Finally, ask a lot of questions during and afterwards; this is where the most valuable data can be found. Also, make sure to take notes and record what the participant is doing and saying during the exercise.

These are the steps I used:

1. *Introduce* the experiment
 - 1a. Encourage them to look for problems
 - 1b. Sign the consent form
2. *Demonstrate* what a think aloud is
3. *Ask* background questions
4. *Present* the tasks
5. *Perform* the tasks (users, not researchers!)
6. *Ask* post-experiment questions
7. *Thank* the participants



A screengrab from one of the think alouds

I had two users conduct think alouds while using some of WordPress' basic features in order to test the UI and organization of the site. My hypotheses were that the users would have significant trouble completing the tasks because the features are nested in a non-intuitive manner and that the users would not have a pleasant experience. Overall, my hypotheses were largely supported by the data: participants had difficulty completing the tasks and reported strong feelings of confusion and frustration.

Based on my findings, I was able to identify key problematic features and assign severity ratings to them. These are the issues that deserve the most attention of the WordPress team:

'Edit' Button – 3

Clicking this button was a critical incident as users spent a considerable amount of time on this page and became increasingly frustrated as they could not find anything seemingly helpful on this page.

'My Site' Menu – 4

This menu cultivates an inaccurate mental model of website structure by giving inordinate access to different features, as one user said.

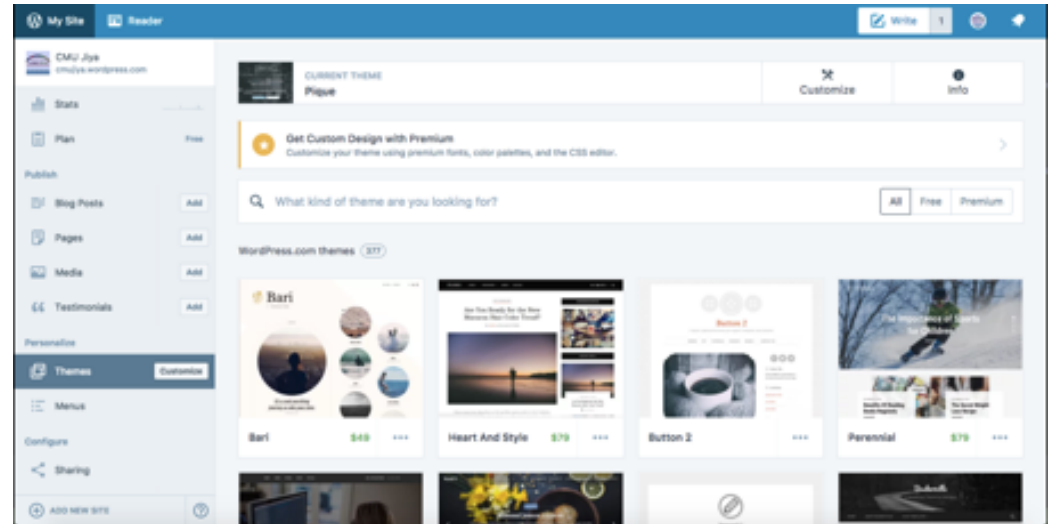
'Themes' – 2

This created some confusing amongst my participants, but they were able to eventually achieve their desired outcome. The focus needs to be shifted to editing the current design rather than completely redesigning the site.



A think aloud participant

Think alouds are a great way of identifying problems in user interfaces and looking at services from the *user's perspective*. I had the added benefit of conducting the study in-person, which gave me more direct access to and allowed for a better rapport with the participants. With the help of think alouds, I was able to *uncover some major flaws* in WordPress' system that should be improved.



Current WordPress UI

A/B Testing

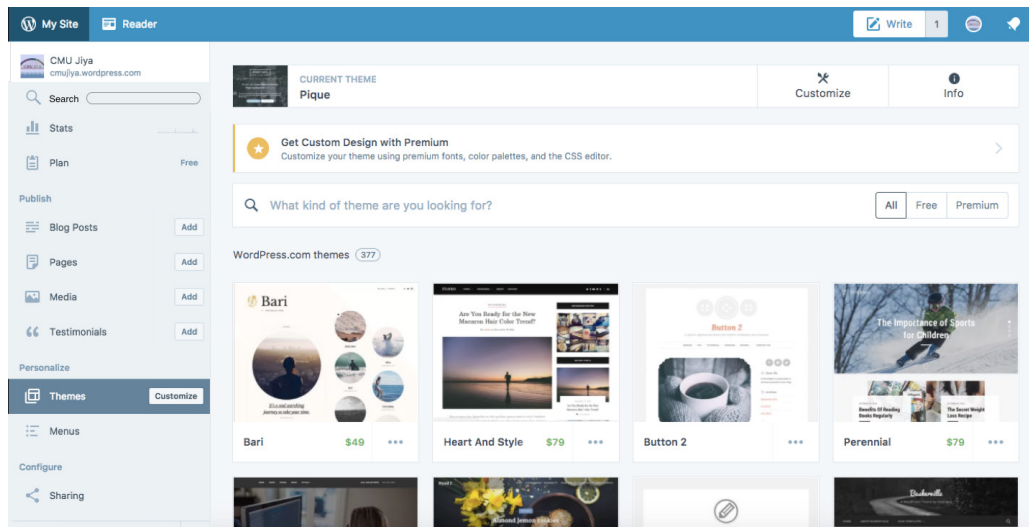
Team: Nila Banerjee

Timeframe: 1 Week

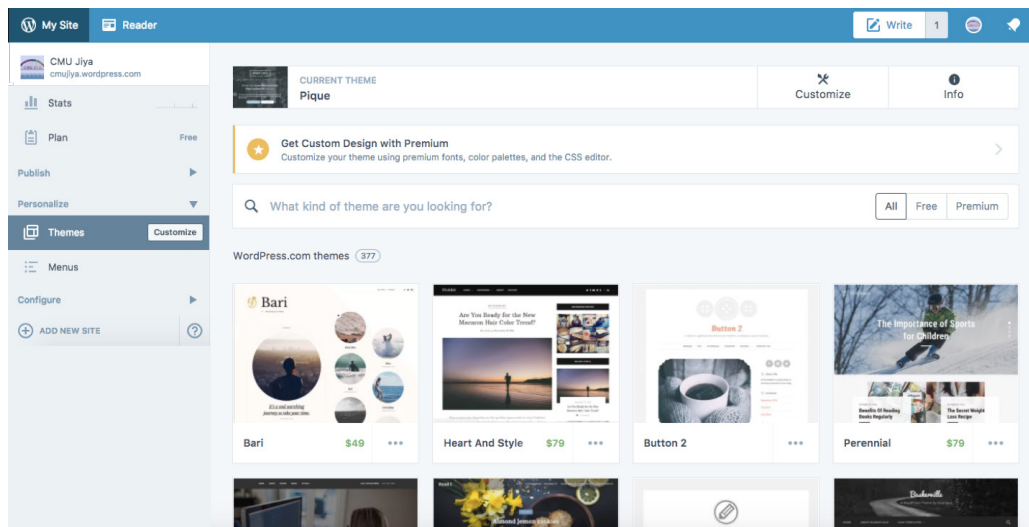
Deliverables: Formal Report with Research Plan

Now that you've identified some problems in your UI, you might be wondering how to make them better. Will the changes you make actually cause any improvements in the user experience? One of trying out different UIs is A/B Testing, in which groups of users are exposed to different UIs in order to compare efficacy. Usually, success is measured using conversion rate, or the proportion of users who performed the desired task, such as purchasing a product or signing up for a mailing list.

A/B testing allows researchers to *monitor user behavior in natural environments and workflows* and provides us with real data to act as a basis for design decisions. Moreover, since the focus of the test is usually narrow (a specific change in workflow, interfaces, etc.), the manipulated variables demonstrate *direct effects* on usability. A/B tests are the best way to determine if a product is more financially successful as conversion and click-through rates are easily measured. Finally, disseminating an updated UI to half (or even a large subset) of one's users ensures a *substantial and representative sample*, whereas other user testing methods rely on findings from only a small (and, therefore, perhaps skewed) demographic.



Version 1



Version 2

However, A/B tests only demonstrate what people do, not why they do it. From the results of such a test, we cannot determine how the user felt or perceived the UI, but rather only for which one they had a demonstrated preference.

For my A/B test, I created *two new versions* of WordPress' 'MySite' menu, which I found to be a focal point of frustration in my think alouds. These versions are shown on the left, but do not possess any functionality. My hypothesis was that there would a difference in the amount of money spent on premium WordPress features between the two versions. In my test, half of WordPress' users would be directed to the first UI and the other half to the second UI. To minimize bias, the users will be randomly directed to either of the UIs; to account for possible differences between days of the week and to gather a large sample size, the test will occur over a period of two weeks. The following data should be collected from all users: the UI they were shown and the amount of money they spent on WordPress during the test period. At the end of the test, the data will be compared to discern if one UI caused a statistically significant increase in conversion. This UI should then be officially integrated into the WordPress system if profits appear to largely outweigh implementation and maintenance costs.

Due to time and resource constraints, I was not actually able to perform the tests. However, I was able to get a feel for A/B testing and the ways in which such a test can be designed. Evidence from past A/B tests has shown that even seemingly *minor changes in UI can greatly impact the number of users willing to spend money on a product*. Testing interfaces before publishing them permanently can save companies significant time and money. A/B testing creates a powerful route to *quantitatively measure the efficacy* of different UIs, although it can leave researchers wondering why one interface is better than another.

Heuristic Evaluation

Team: Bobbie Soques, Divya Mohan, Lois Kim, Nila Banerjee, Steven Ji

Timeframe: 1 Week

Deliverables: Formal Report with Findings

Besides users, experts can also be a significant source of evaluation, for example, through heuristic evaluations. A heuristic evaluation is conducted by an *UI expert* who gauges the extent to which certain pre-determined heuristics are met by a product or service. Generally, the heuristics are *overarching themes* that can be applied to a variety of interfaces. One of the most popularly implemented (and the one used in this project) is Nielsen's 10 Heuristics for UI Design, which includes items like error prevention and user control and freedom. These tests allow for rapid feedback with *pinpointed problem areas* and can be logistically smoother than recruiting users. Yet, these evaluations are clouded by a degree of expert blindness and the experts used don't necessarily represent the average user.



To perform a heuristic evaluation, find and familiarize yourself with a list of heuristics that you feel should be exemplified in the product. While common lists are great, it can sometimes be necessary to modify them to get a better list of ideals for your product. Then, select evaluators who will analyze the product based on the provided heuristics. Research has shown that just *5 evaluators can capture the large majority of usability problems*, so it often isn't necessary to recruit a large number of people to reap benefits. Looking through the problems discovered by evaluators can shed light on specific usability problems, as well as *patterns of breakdowns*.

In this project, a team of 5 evaluators, also known as Team Shrimp, analyzed Duolingo, a platform that allows users to learn new languages. With over 150 million users, Duolingo is a tool for students (and sometimes educators) to capture their progress towards a new language with gamified lessons. We first assigned each evaluator a persona with which the website should be approached to gain a wide range of perspectives and tasks executed on the site. We then went through the site and evaluated it based on Nielsen's heuristics. My persona was based on a college student interested in studying abroad. I created a number of Usability Aspect Reports (UAR) based on the problems I found in Duolingo's site.



Jasmine Amari
College Student
Age 19

"I want to be able to communicate with locals and really understand Swiss culture"

Perception of Duolingo
Thinks Duolingo will brush up her knowledge of French and help with conversational idiosyncracies.

Pain Point
Hasn't really spoken French since high school courses.

Short Term Goal
To be able to communicate effectively with Swiss residents.

Long Term Goal
1. Use language to better connect with foreign cultures.
2. Be fluent in conversational French.

Impersonated by
Nila

Use cases
1. Register for an account
2. Go through one lesson
3. searched up tips on the past tense in the discussion forum and in lessons

Personality
Driven, adventurous, studious, unique, popular, caring.

Motivation
Wants to travel the world and deeply connect with people and cultures, hopes to be a business analyst and work with people from diverse backgrounds.

Persona I used

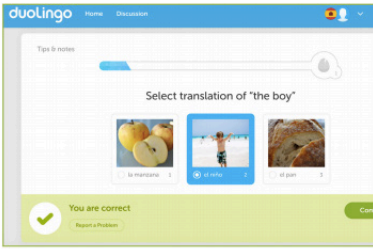
Highlighted UARs + Possible Solutions

Common Errors Mid-lesson

Several researchers encountered errors mid-lesson, such as being unable to follow the "Discuss sentence" link or play recordings of words. Solutions to these errors could only be found by exiting the lesson to search the documentation or refreshing the page, thus causing the user to lose their progress in the lesson. The research team considers this problem to be high-impact, given that errors like these are common across multiple users, can happen multiple times per lesson, and cause considerable frustration to the user.

Possible Solutions:

- Provide brief FAQs/Troubleshooting information on the lesson page, so that the user can find solutions to common problems without exiting the lesson
- Save the user's progress in a lesson so they can safely exit and return
- Recognize when an error happens and provide a diagnosis and solution in a pop-up



There is no FAQ/help readily available when the user mid-lesson.

Example UAR



Team Shrimp evaluating heuristic problems

These are the overarching problems we detected:

1. The *on-boarding process* is confusing for the user and does not present all of the options equivalently. Users can not sign up as educators or complete their profile in one fluid step, disrupting their workflow.
2. The website presents multiple *dark patterns*, such as advertisements for languages that are not available to learn, which can cause users to be misguided into setting up an account that ultimately dissatisfies them.
3. There is a consistent *lack of effective feedback* throughout the website, leaving the user confused as to both the system's status and their progress in their lessons.
4. During lessons, many users seem to experience the *same errors*, such as problems with pop-ups, of which the site is aware, so they should present aid. However, there is a lack of recognition and navigation through these errors, forcing users to either deal with the error or dig through documentation for solutions.
5. A significant amount of Duolingo's *content is hidden* from the user at first glance, including educator sign up and several settings. This makes it quite difficult for users to quickly and easily access relevant content, sometimes giving off the impression that those features simply do not exist.

Heuristic evaluations help *provide an expert eye* on products and can elucidate patterns or problem areas that may have simply been accepted by users. Nonetheless, it is important to keep in mind that *experts are not users* and that the best researchers would look to both groups for feedback.

PART III:

UCRE & ME

One of the things we worked on in this class was developing mission statements, a sentence that encompasses the focus and goals of a project. This is my personal mission statement:

*To do work that uplifts
and empowers women
from all backgrounds*

I have always been passionate about women-oriented issues and wanted to work towards a future in which women are truly seen and treated as equals. While the current sociopolitical climate is discouraging, I believe and hope that women, as a group, will make significant strides in the next few decades. In the past, the feminist movement has largely been focused on catering to specific demographics of the female population; in the future, I hope that the movement can grow and be more inclusive of all types of women. It is my personal goal to become more deeply involved in this fight and work towards making the world a better place for everyone by creating more opportunities for and strengthening the position of women.

Over the semester, I have really enjoyed taking UCRE and have learned a lot more than I could have anticipated. Even though I've only been working on class projects, I feel like I have a degree of industry experience that will become a major strength for me in my upcoming summer internship working on UX Research. I enjoyed working on projects dealing with real products and issues, although the workload could be overwhelming at times. Some weeks, especially in the first half of the semester, I felt like I was only doing homework for UCRE. I think it would be helpful if some of the projects were broken down into smaller chunks or if we had more time to work them. Furthermore, while I appreciate the intention behind asking teams to research problems related to Pittsburgh, I feel that it made the project needlessly difficult. Not only was finding people to interview a logistical nightmare, I felt rude asking them to discuss their personal issues with me for just a class project. It would have been better if the project was focused on problems at Carnegie Mellon. All in all, I had a great semester in UCRE!